Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators

M. Hagmann[a] O. Scaillet[b]

First Version: September 2003

Abstract

We consider semiparametric asymmetric kernel density estimators when the unknown density has support on $[0, \infty)$. We provide a unifying framework which contains asymmetric kernel versions of several semiparametric density estimators considered previously in the literature. This framework allows us to use popular parametric models in a nonparametric fashion and yields estimators which are robust to misspecification. We further develop a specification test to determine if a density belongs to a particular parametric family. The proposed estimators outperform rival non- and semiparametric estimators in finite samples and are simple to implement. We provide applications to loss data from a large Swiss health insurer and Brazilian income data.

Key words and phrases: semiparametric density estimation, asymmetric kernel, income distribution, loss distribution, health insurance, specification testing.

JEL Classification: C13, C14.

[a] HEC Lausanne and FAME, BFSH 1, CH-1015 Lausanne-Dorigny.
E-mail: matthias.hagmann@hec.unil.ch.
[b] HEC Genève and FAME, UNIMAIL, 102 Bd Carl Vogt, CH-1211 Genève.
E-mail: olivier.scaillet@hec.unige.ch.

1 Introduction

One of the major concerns of insurance companies is the study of a group of risks. For insurers, a good understanding of the size of a single claim is of most importance. Loss distributions describe the probability distribution of a payment to the insured. Traditional methods in the actuarial literature use parametric specifications to model single claims. The most popular specifications are the lognormal, Weibull and Pareto distributions. Hogg and Klugman (1984) and Klugman, Panjer and Willmot (1998) describe a set of continuous parametric distributions which can be used for modelling a single claim size. It is, however, unlikely that something as complex as the generating process of insurance claims can be described by just a few parameters. A wrong parametric specification may lead to an inadequate measurement of the risk contained in the insurance portfolio and consequently to a mispricing of insurance contracts.

In a totally different area or research, economists studying income distributions and income inequality use similar parametric models to estimate the distribution of income and its evolution over time. Popular models are the gamma, lognormal and Pareto distributions, see Cowell (1999). Whereas the lognormal distribution is thought to have the best overall shape, the Pareto is considered to be a more suitable distribution for individuals in the upper end of the income distribution. Although these densities may capture some stylized facts of income distributions, it is again unlikely that income distributions can be described by just a few parameters. The imposition of a wrong parametric model may lead to inconsistent estimates and misleading inference, as well as to disputable conclusions in inequality measurement for example.

A method which does not require the specification of a parametric model is nonparametric kernel smoothing. This method provides valid inference under a much broader class of structures than those imposed by parametric models. Unfortunately, this robustness comes at a price. The convergence rate of nonparametric estimators is slower than the parametric rate and the bias induced by the smoothing procedure can be substantial even for moderate sample sizes. Since both income and losses are positive variables, the standard kernel estimator proposed by Rosenblatt (1956) has a boundary bias which is due to weight allocation by the fixed symmetric kernel outside the support of f when smoothing close to the boundary is carried out. Additionally, standard kernel methods often fail to provide good estimation in the tails of the distribution. This is of particular importance for the estimation of loss distributions to get appropriate and precise risk measures when designing an efficient risk management system.

We propose a semiparametric estimation framework for the estimation of densities which have support on $[0, \infty)$. Our estimation procedure can deal with all problems of the standard kernel estimator mentioned previously, and this in a single way. Although the above parametric models may be inaccurate, they can be used in a nonparametric fashion to help to decrease the bias induced by nonparametric smoothing. If the parametric model is accurate, the performance of our semiparametric estimator can be close to pure parametric estimation. Following Hjort and Glad (1995) (H&G), we start with a parametric estimator of the unknown density (economic theory may help in providing the parametric model), and then correct nonparametrically for possible misspecification. To decrease the bias even further, we give some local parametric guidance to this nonparametric correction in the spirit of Hjort and Jones (1996) (H&J). We call this approach *local multiplicative bias correction*, or short LMBC.

To address the boundary bias problem induced by the symmetric kernel, we develop LMBC in an asymmetric kernel framework. Asymmetric kernel estimators were recently proposed by Chen (2000) as a convenient way to solve the boundary bias problem. The symmetric kernel is replaced by an asymmetric gamma kernel which matches the support of the unknown density[1]. As an alternative to the gamma kernel, Scaillet (2003) introduced kernels based on the inverse Gaussian and reciprocal inverse Gaussian density. All of these kernel functions have flexible form, are located on the nonnegative real line and produce nonnegative density estimates. Also, they change the amount of smoothing in a natural way as one moves away from the boundary. This is particularly attractive when estimating densities which have areas sparse in data because more data points can be pooled. As pointed out by Cowell (1999) "Empirical income distributions typically have long tails with sparse data". The same holds true for empirical loss distributions and we therefore think that these kernels are very well suited in this context. The variance advantage of the asymmetric kernel comes, however, at the cost of a slightly increased bias as one moves away from the boundary compared to symmetric kernels, which highlights the importance of effective bias reduction techniques in this area of research. In a vast simulation study, Scaillet (2003) obtains attractive finite sample performance of these asymmetric kernel estimators. He also reports that boundary kernel estimators lead very often to negative density estimates without outperforming asymmetric kernel density estimators. Chen (2000) reports superior performance of the gamma kernel estimator compared to other remedies proposed in the literature as the local linear estimator of Jones (1993). A particular advantage of the gamma kernel estimator is its consistency when the true density is unbounded at $x = 0$, which is important for the estimation of highly skewed loss and income distributions. This is shown in Bouezmarni and Scaillet (2003) who also establish uniform and L_1 convergence results for asymmetric kernel density estimators. Furthermore they report nice finite sample performance of the asymmetric kernel density estimator w.r.t. the L_1 norm.

In a simulation study, we find that LMBC in connection with asymmetric kernels yields excellent results. These estimators perform better in a mean integrated squared error (MISE) sense than pure nonparametric estimators. If the parametric information provided is accurate, we find that a MISE

[1] Other remedies include the use of particular boundary kernels or bandwidths, see e.g. Rice (1984), Schuster (1985), Jones (1993), Müller (1991) and Jones & Foster (1996).

reduction of 50-80% can be reasonably expected. Even if the misspecification considered is large, our best estimator still achieves a MISE reduction of around 25%. Asymmetric kernel based LMBC estimators also outperform their symmetric rivals thanks to their intrinsic advantage in the tails of the density. Furthermore, they are often simpler to implement.

As a by-product of our approach, we propose an attractive semiparametric specification test to determine whether a particular unknown density belongs to a parametric class of densities. We also show how this statistic can be used to determine which density can be felt as a suitable parametric start.

Although we concentrate in the sequel on loss and income distributions, similar issues as discussed above are also important in the finance literature. Aït-Sahalia (1996a, 1996b) develops an estimation and specification testing procedure for diffusion models of the short term interest rate. In this framework, the nonparametric estimation of the stationary distribution of the interest rate process plays a key role. The intertemporal general equilibrium asset pricing model of Cox, Ingersoll and Ross (1985) implies that this distribution follows a gamma probability law. Although this interest rate model may again be overly restrictive, it gives some economic guidance about the likely form of the stationary distribution of the short rate. This information can be incorporated in a semiparametric estimator like ours. Furthermore, our results are potentially important for estimation and specification testing of the baseline hazard function in autoregressive conditional duration (ACD) models. In this literature parametric models like the Burr and generalized gamma distribution are popular specifications for the baseline hazard. We refer to Engle (2000) for an overview, and Fernandes and Grammig (2000) for exploitation of asymmetric kernels in financial duration analysis. Clearly the standard kernel estimator is again not appropriate in these contexts, since it does not take into account that the underlying variables, interest rates and durations, are nonnegative.

The outline of the paper is as follows. In Section 2 we introduce our semiparametric estimation framework and relate it to the relevant associated literature. This unified framework embeds semiparametric density estimators developed by H&G, H&J and Loader (1996) and allows us to clarify the interdependence between these approaches. Section 3 recalls asymmetric kernel methods. Section 4 contains the main contribution of the paper, namely the extension of the LMBC framework to the asymmetric kernel case. We develop several examples, which show that the estimation procedure is user friendly and remarkably simple to implement in most cases[2]. The procedure is therefore appealing for applied work. We also discuss bandwidth choice and model diagnostic tests. In Section 5 we compare the performance of our estimators through an extensive simulation study. To the best of our knowledge, it is the first time that the various semiparametric approaches mentioned above are compared on a finite sample basis. In Section 6 we provide two empirical applications: the first one to loss data from a

[2]MATLAB code can be freely downloaded from http://www.unibas.ch/wwz/makro/

large Swiss health insurer, the second one to Brazilian income data. Section 7 contains some concluding remarks. An appendix gathers the proofs and technicalities related to the properties of the various estimators considered in the text.

2 Local multiplicative bias correction

In nonparametric regression, local polynomial fitting is a very popular approach, e.g. Ruppert and Wand (1994) and the references therein. Gozalo and Linton (2000) were the first to consider local fitting of a general functional using a least squares criterion, a normal error distribution version of a local likelihood estimator. For density estimation, the local likelihood approach was independently developed at the same time by H&J and Loader (1996). For a recent extension of the approach with application to Value-at-Risk (VaR) in risk management, see Gourieroux and Jasiak (2001). Whereas Loader (1996) concentrates on local polynomial fitting to the logarithm of the density, H&J allow for general local functionals like Gozalo and Linton (2000) in regression estimation.

In this section, we briefly introduce an estimation framework based on familiar symmetric kernel methods, which contains as special cases the local likelihood and multiplicative bias correction approach as described in H&J, Loader (1996) and H&G, respectively. This framework will allow us to derive the properties of these methods using asymmetric kernels in a single step, instead of treating the methods separately. In addition, this framework allows one to shed some light on the interdependence between these methods.

Let $X_1, ...X_n$ be a random sample from a probability distribution F with an unknown density function f which has support on $[0, \infty)$. Furthermore we choose a local model for f given by

$$m\left(x, \theta_1, \theta_2\left(x\right)\right) = f\left(x, \theta_1\right) r\left(x, \theta_2\left(x\right)\right), \tag{2.1}$$

where $f(x, \theta_1)$ is a parametric family of densities indexed by the parameter $\theta_1 = (\theta_{11}, ..., \theta_{1p}) \in \Theta_1 \in \mathbf{R}^p$, $r\left(x, \theta_2\left(x\right)\right)$ is a model for the correction function $r(x) = f(x)/f(x, \theta_1)$ where $\theta_2\left(x\right) = (\theta_{21}, ..., \theta_{2q}) \in \Theta_2 \in \mathbf{R}^q$. H&J treat this local model as a particular example in their paper, whereas we use it to embed several estimators known in the literature. We call this local multiplicative bias correction (LMBC) since only the multiplicative correction factor is modelled locally. The procedure is as follows: first, the parameter θ_1, which does not depend on x, is estimated by maximum likelihood. It is well known that when the parametric model $f\left(x, \theta_1\right)$ is misspecified, θ_1 converges in probability to the pseudo true value θ_1^0 which minimizes the Kullback-Leibler distance of $f\left(x, \theta_1\right)$ from the true $f\left(x\right)$, see e.g. White (1982) and Gourieroux, Monfort and Trognon (1984). Second, choose $\theta_2\left(x\right)$ such that

$$\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)v(x, X_i, \theta_2) - \int K_h(t-x)v(x, t, \theta_2) f\left(t, \hat{\theta}_1\right) r\left(t, \theta_2\right) dt = 0 \tag{2.2}$$

holds, where $K_h(z) = (1/h)K(z)$ is a symmetric kernel function, h is the bandwidth parameter and $v(x,t,\theta_2)$ is a $q \times 1$ vector of weighting functions. We omit for notational simplicity the possible dependence of the weighting function on θ_1. If we choose the score $\partial \log m(x,\theta_1,\theta_2(x))/\partial\theta_2$ as the weighting function, then Equation (2.2) is just the first order condition of the local likelihood function given in H&J. The local multiplicatively bias corrected density estimator is

$$\hat{f}(x) = f\left(x,\hat{\theta}_1\right) r\left(x,\hat{\theta}_2(x)\right). \tag{2.3}$$

Since $\hat{\theta}_1$ exhibits $\sqrt{n}$-convergence which is faster than the nonparametric rate, the additional variability introduced through the first step estimation of θ_1 does not influence the bias and variance of $\hat{f}(x)$ up to negligible higher order terms. We give a more formal discussion in the asymmetric kernel case in Section 4. From the theoretical results concerning bias and variance of the local likelihood estimator given in H&J, it immediately follows that this estimator has the same variance as the standard kernel estimator introduced by Rosenblatt (1956). The bias is however different. Compared to H&J in their treatment of this example, we do express the bias in terms of the correction factor. This is more intuitive and allows a better comparison between different estimation strategies. When θ_2 is one dimensional, the bias is

$$Bias\left(\hat{f}(x)\right) = \sigma_K^2 h^2 \left(\begin{array}{c} \frac{1}{2} f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right] \\ +\{\frac{v_0^{(1)}(x)}{v_0(x)} f_0(x) + f_0^{(1)}(x)\}\left[r_0^{(1)}(x) - r^{(1)}(x,\theta_2^0)\right] \end{array} \right), \tag{2.4}$$

where $\sigma_K^2 = \int K(z)^2 dz$ and the subscript 0 denotes substitution of the pseudo true value for θ_1. The magnitude of this term depends on how well the correction function can be approximated locally by a suitable parametric model. This is so if $r(x)$ is smooth, or equivalently, if the global parametric start is close to the true density. In the single parameter case, the bias also depends on the weighting function and on the distance between the slopes of the correction function and its local model. If $\dim(\theta_2) \geq 2$, the bias is free of this term and only the first term in the brackets appears. For further details we refer to H&J.

Direct local modelling of the density can be obtained by choosing the parametric start density as an improper uniform distribution. W.l.o.g. set $f_0(x)$ to one. Then the only source of bias reduction is provided by the local model $r(t,\theta)$. The bias is given by $\frac{1}{2}\sigma_K^2 h^2\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]$ as in H&J.

The multiplicatively corrected kernel estimator of H&G emerges from choosing the weighting function as $1/f\left(x,\hat{\theta}_1\right)$ and choosing the local model as a constant. From Equations (2.2) and (2.3) it follows that the estimator is

$$\hat{f}(x) = f\left(x,\hat{\theta}_1\right)\frac{1}{nh}\sum_{i=1}^{n} K_h(X_i - x)\frac{1}{f\left(X_i,\hat{\theta}_1\right)} / \int K_h(t-x)dt. \tag{2.5}$$

From (2.4), the bias is $(1/2)\,\sigma_K^2 h^2 f_0(x)\, r^{(2)}(x)$, which is the bias obtained by H&G and does not depend on the chosen weighting function. This is the only possible choice of weighting function which sets the second bracket term in Equation (2.4) to zero. In Equation (2.5), the term in the denominator is one in the interior. It does, however, provide close to the boundary a correction for the fact that the symmetric kernel is allocating weight outside the support of the density. This correction is not optimal and boundary bias is still of the undesirable order $O(h)$. Like in nonparametric regression, see e.g. Fan and Gijbels (1992), one of the possible boundary bias correction methods which achieves an $O(h^2)$ order is the popular local linear estimator, see Jones (1993) for the density case. To obtain a local linear H&G version we choose the local model as $r(t,\theta_2) = \theta_{21} + \theta_{22}(t-x)$ and the weight functions as $1/f(t,\theta_1)$ and $(t-x)/f(t,\theta_1)$. Assuming that the K has support $[-1,1]$[3], the resulting estimator is equivalent to the H&G estimator in the interior of the density. Close to the boundary where $x/h \to \kappa$ and $0 \leq \kappa < 1$, it provides however again a correction due to weight allocation of the symmetric kernel to the negative part of the real line. Define $\alpha_j(\kappa) = \int_{-1}^{\kappa} K(u) u^j du$, then the local linear estimator in the boundary region is

$$\hat{f}(x) = f\left(x, \hat{\theta}_1\right) \frac{\hat{r}(x) - [\alpha_1(\kappa)/h\alpha_2(\kappa)]\,\hat{g}(x)}{\left(\alpha_0(\kappa) - \alpha_1(\kappa)^2/\alpha_2(\kappa)\right)}, \tag{2.6}$$

where $\hat{g}(x)$ is the sample average of $K_h(X_i - x)$. We emphasize that appropriate boundary bias correction is more important in a semiparametric than a pure nonparametric setting. This is because the bias reduction achieved by semiparametric techniques allows to increase the bandwidth and thus to pool more data. This, however, increases the boundary region where the symmetric kernel allocates weight to the negative part of the real line.

After presenting this slight extension of previously proposed estimators, we now turn to an asymmetric kernel version of the above approach. Since the support of these kernels matches the support of the density under consideration, no boundary correction is necessary. We first briefly review asymmetric kernel estimators for densities defined on the nonnegative real line introduced by Chen (2000) and Scaillet (2003). In Section 4 we will treat the LMBC case.

3 Asymmetric kernel methods

The asymmetric kernel density estimator is

$$\hat{f}_b(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i), \tag{3.1}$$

[3] This setup can easily be extended to infinite support kernels. However, finite support is a standard assumption, delineating boundary and interior regions.

where b is a smoothing parameter satisfying $b \to 0$ and $bn \to \infty$ as $n \to \infty$. The asymmetric weighting function K_A is either a gamma density K_{G_1} with parameters $(x/b+1, b)$ as proposed by Chen (2000), an inverse Gaussian density K_{IG} with parameters $(x, 1/b)$ or a reciprocal inverse Gaussian density K_{RIG} with parameters $(1/(x-b), 1/b)$ as proposed by Scaillet (2003). For the related estimation question of densities with compact support, beta kernels can be used, see Chen (1999). These kernel densities K_A are

$$
\begin{aligned}
K^{G_1}_{(x/b+1,b)}(t) &= \frac{t^{x/b}\exp(-t/b)}{\Gamma(x/b+1)\,b^{x/b+1}} I(t \geq 0), \\
K^{IG}_{(x,1/b)}(t) &= \frac{1}{\sqrt{2\pi b t^3}} \exp\left(-\frac{1}{2bx}\left(\frac{t}{x} - 2 + \frac{x}{t}\right)\right) I(t \geq 0), \\
K^{RIG}_{\left(\frac{1}{x-b},1/b\right)}(t) &= \frac{1}{\sqrt{2\pi b t}} \exp\left(-\frac{x-b}{2b}\left(\frac{t}{x-b} - 2 + \frac{x-b}{t}\right)\right) I(t \geq 0).
\end{aligned}
$$

Figure 1 displays the first gamma kernel for some selected x-values. All asymmetric kernels share the property that the shape of the kernel changes according to the value of x. This varying kernel shape changes the amount of smoothing applied by the asymmetric kernel since the variance of for instance $K^{G_1}_{(x/b+1,b)}$ is $xb + b^2$, which is increasing as we move away from the boundary. This is also reflected in the bias and variance expressions, which we give here for the gamma kernel estimator and on which we concentrate in the sequel:

$$
\begin{aligned}
Bias\left(\hat{f}^{G_1}_b(x)\right) &= \left\{f^{(1)}(x) + \frac{1}{2} x f^{(2)}(x)\right\} b + o(b), \qquad (3.2) \\
Var\left(\hat{f}^{G_1}_b(x)\right) &= \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1/2} f(x) & \text{if } x/b \to \infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)} n^{-1} b^{-1} f(x) & \text{if } x/b \to \kappa. \end{cases}
\end{aligned}
$$

This estimator is not subject to boundary bias, but involves the first derivative of the unknown density. This is because x is not the mean of the gamma kernel $K^G_{(x/b+1,b)}$, rather its mode. This is different for the inverse Gaussian and reciprocal inverse Gaussian kernel estimators, whose biases only involve the second order derivative of the unknown density. To circumvent the first derivative in the bias expression, Chen (2000) proposes the use of a different gamma kernel $K^{G_2}_{(\rho_b(x),b)}$ where $\rho_b(x)$ is

$$
\rho_b(x) = \begin{cases} x/b & \text{if } x \geq 2b, \\ \frac{1}{4}(x/b)^2 + 1 & \text{if } x \in [0, 2b). \end{cases}
$$

This second gamma kernel estimator has

$$
Bias\left(\hat{f}^{G_2}_b(x)\right) = \begin{cases} \frac{1}{2} x f^{(2)}(x) b + o(b) & \text{if } x \geq 2b, \\ \xi_b(x) b f^{(1)}(x) + o(b) & \text{if } x \in [0, 2b), \end{cases} \qquad (3.3)
$$

where $\xi_b(x)$ takes the form $(\rho_b(x) - x/b)$. For the second gamma kernel, the first derivative appears in the bias only in the boundary. This is, however, compensated by the disappearance of $f^{(2)}(x)$. Scaillet

(2003) reports that the RIG kernel has a similar shape as the second gamma kernel but has a slightly better finite sample performance evaluated in a simulation study. The variance expression of the second gamma kernel is close to that of the first one and therefore not reported here.

Compared to other boundary correction techniques, the bias of gamma kernel estimators may be larger as x increases, this can however be compensated by a reduced variance. The variance of the gamma kernel estimator decreases as x gets larger which is in contrast to symmetric kernel estimators whose variance coefficients remain constant outside the boundary area. Also asymmetric kernels have a larger effective sample size than kernels with compact support. This is desirable for estimating densities with sparse areas as more data points can be pooled.

In the following we address the question of semiparametric bias reduction techniques for asymmetric kernel methods. This is important since as reported, the bias may be larger than for standard symmetric kernel methods. Effective bias reduction techniques combined with the fact of a decreasing variance as we move away from the boundary is giving us hope for promising performance of our estimators for the estimation of loss and income distributions. This will be confirmed later in the paper.

4 Local multiplicative bias correction with asymmetric kernels

Apart from being an attractive semiparametric bias reduction framework, LMBC allows us to implement a popular boundary bias reduction by choosing a local linear model for the density or correction factor. This boundary bias reduction is not necessary in the asymmetric kernel framework since no weight is allocated outside the support of the unknown density. The effect of LMBC in an asymmetric framework is just to reduce the potentially larger bias for asymmetric kernel techniques. In addition to the attractive feature of asymmetric kernels mentioned in the introduction, a further advantage is that despite the seemingly complexity of the approach presented below, asymmetric kernels allow the construction of user friendly estimators. Symmetric kernels like e.g. the Gaussian kernel often require numerical integration and optimization procedures.

We now extend the LMBC approach to the asymmetric kernel case, compute bias and variance of the estimator and discuss the choice of bandwidth. We also consider the special cases of H&J, Loader (1996) and H&G and show how these methods can be applied to the estimation of income and loss distributions.

4.1 Definition of the estimator

We follow the notation introduced in Section 2. The estimator is $\tilde{f}_A(x) = f\left(x, \hat{\theta}_1\right) r\left(x, \hat{\theta}_2(x)\right)$, where $\hat{\theta}_1$ is the global maximum likelihood estimator which does not depend on x and $\hat{\theta}_2(x)$ is chosen by maximizing the local likelihood function

$$\begin{aligned} \log L_n(x,\theta) &= \int_0^\infty K_A(x,b)(t)\left\{\log m\left(t,\hat{\theta}_1,\theta_2\right) dF_n(t) - m\left(t,\hat{\theta}_1,\theta_2\right) dt\right\} \\ &= \frac{1}{n}\sum_{i=1}^n K_A(x,b)(X_i)\log m(X_i,\hat{\theta}_1,\theta_2) - \int_0^\infty K_A(x,b)(t)\, m(t,\hat{\theta}_1,\theta_2)dt, \end{aligned} \tag{4.1}$$

with F_n denoting the empirical distribution function. This criterion function is equivalent to the one of H&J. However, the symmetric kernel is replaced by an asymmetric kernel, whose support matches the support of the density we wish to estimate. For notational simplicity we omit the local dependency of θ_2 on x. The first term in (4.1) is the standard log-likelihood function weighted by an asymmetric kernel function. Maximizing this term alone would lead to inconsistent results because the expectation of its score is not equal to zero at the true parameter value θ_2^0. The second term guarantees that this is the case, we refer to H&J. From (4.1), when b is very large, $K_A(x,b)(t)$ is independent of t and the above expression is a constant times the ordinary, normalized log-likelihood function. The maximization of the local likelihood then becomes the same as ordinary likelihood maximization. But if b is small, the maximization of $\log L_n(x,\theta)$ will provide the best local estimator of $f(x)$. This follows since under some regularity conditions,

$$\log L_n(x,\theta) \to_p \pi(x,\theta) = \int_0^\infty K_A(x,b)(t)\left\{f(t)\log m(t,\theta_1^0,\theta_2) - m(t,\theta_1^0,\theta_2)\right\}dt,$$

as n grows. Hence $\hat{\theta}_2$, the maximizer of (4.1), aims at the parameter value $\theta_2^0(x)$ that maximizes $\pi(x,\theta)$ when the above convergence is uniform over the parameter space of θ_2. See for example Linton and Pakes (2001). The solution to the above problem minimizes the following distance measure which is a localized version of the Kullback-Leibler distance of $f(x)$ from $m(x,\theta)$:

$$d[f, m(\cdot,\theta)] = \int_0^\infty K_A(x,b)(t)\left[f(t)\log\frac{f(t)}{m(t,\theta)} - \{f(t) - m(t,\theta)\}\right]dt.$$

This shows that $\hat{\theta}_2$ aims at the best local parametric approximant to the true f. The estimator depends on the chosen smoothing parameter. For further details and justifications of the local likelihood approach, see H&J, Loader (1996) and the references therein.

The estimator $\hat{\theta}_2$ for general weight functions $v(x,t,\theta_2)$ is defined to be the solution to

$$\frac{1}{n}\sum_{i=1}^n K_A(x,b)(X_i)\, v(x,X_i,\theta_2) - \int_0^\infty K_A(x,b)(t)\, v(x,t,\theta_2)m(t,\hat{\theta}_1,\theta_2)dt = 0. \tag{4.2}$$

This is identical to the first order condition of (4.1) in the case where v is chosen as the score $u(t,\theta_2) = (\partial/\partial\theta_2)\log r(t,\theta_2)$. For identification reasons, assume that

$$V_n(x,\theta_2) \to^p V(x,\theta_2) = \int K_A(x,b)(t)\, v(x,t,\theta_2) f_0(x)\{r(t) - r(t,\theta_2)\}dt = 0$$

has a unique solution $\theta_2 = \theta_2^0$. This requires that the q weight functions are functionally independent and that the correction function $r(t)$ is within reach of the parametric model $r(t,\theta_2)$ as θ_2 varies. This is like M-estimation in a possibly misspecified case, since the true $r(x)$ does not have to belong to the parametric family $r(t,\theta)$.

4.2 Large sample properties

We now develop the bias and variance of the LMBC estimator. The derivations of all results presented here are given in the appendix. When not stated otherwise, we will focus on results for the first gamma kernel developed in Chen (2000) since other kernel choices can be handled in a similar fashion.

If we locally fit one parameter, the bias of the LMBC estimator is

$$\begin{aligned} B\left(\tilde{f}_{G_1}(x)\right) &= f_0(x)\left[\{r_0^{(1)}(x) - r^{(1)}(x,\theta_2^0)\} + \frac{1}{2}x\{r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\}\right] b \\ &+ \left(\frac{v_0^{(1)}(x)}{v_0(x)} f_0(x) + f_0^{(1)}(x)\right) x\{r_0^{(1)}(x) - r^{(1)}(x,\theta_2^0)\} b \\ &+ o(b) + O\left(\frac{1}{nb^{1/2}}\right), \end{aligned} \tag{4.3}$$

where we use the same notation as in the second section of this paper. For the second gamma kernel proposed by Chen (2000), the first derivative term in the square brackets would not appear in the bias in the interior. H&J also note that the first derivative will vanish automatically from the bias expression for $q \geq 2$. Equation (4.3) will then hold for any component $v_{j,0}(x)$ of the weighting function, which can only be true if $\left[r_0^{(1)}(x) - r^{(1)}(x,)\right] = o(1)$ as $b \to 0$. This is not generally the case with one parameter. This means that for $q \geq 2$, the bias reduces to

$$Bias\left(\tilde{f}_{G_*}(x)\right) = \frac{1}{2} f_0(x)\{r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\} x b + o(b) + O\left(\frac{1}{nb^{1/2}}\right), \tag{4.4}$$

independent of whether we use the first or second gamma kernel. H&J show that for $q \geq 3$ one can argue that $\{r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\}$ is also $o(1)$ and third and fourth order derivatives appear in the bias term. This property also holds for asymmetric kernels.

There are several worthwhile remarks. First, we obtain the same result as in the symmetric kernel case. Comparing Equations (3.3) and (4.4), the second derivative in the bias of the asymmetric kernel estimator is replaced by $f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]$. So this estimator performs better than pure asymmetric kernel methods if the latter expression is smaller than the former. This is the case if the parametric start is close to the true density since then $r_0^{(2)}(x)$ is small. Additionally, the local model for the correction factor can make this term even smaller if it can locally capture the curvature of the correction factor. If the model is correct, the local likelihood estimator is unbiased up to the order considered.

The variance of the asymmetric LMBC estimator in the one parameter case is

$$Var\left(\tilde{f}_b^{G_1}(x)\right) = \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1/2} f(x) & \text{if } x/b \to \infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)} n^{-1} b^{-1} f(x) & \text{if } x/b \to \kappa, \end{cases} \tag{4.5}$$

where κ is a positive constant. We therefore obtain exactly the same result for the variance as for the pure nonparametric gamma kernel estimator. Also the variance of the LMBC estimator does not depend on the chosen weighting functions. We therefore have some flexibility to choose them to obtain estimators which are tractable to implement.

The variance of the LMBC estimator in the multiple parameter case for a general asymmetric kernel is

$$Var\left(\tilde{f}_A(x)\right) = \frac{f(x)}{n} e_1^t \tau\left(K_A\right) e_1 - \frac{f(x)^2}{n} + O\left(b/n\right), \tag{4.6}$$

where, using some simplified notation, $\tau\left(K_A\right)$ is given by

$$\left(\int_0^\infty K_A(t) V_t V_t' dt\right)^{-1} \left(\int_0^\infty K_A(t)^2 V_t V_t' dt\right) \left(\int_0^\infty K_A(t) V_t V_t' dt\right)^{-1}$$

and V_t is $q \times 1$ vector containing in the j^{th} position the elements $(t-x)^{j-1}$ for $j = 1, ..., q$. This expression depends on the kernel being used. Independent of the kernel used, the variance of the LMBC estimator in the two parameter case is the same as in the single parameter case. We collect results for all the asymmetric kernel estimators in the following proposition.

Proposition 1 *The bias expressions of the asymmetric LMBC estimator in the cases where K_A is the first or second gamma, inverse Gaussian and reciprocal inverse Gaussian kernel are given for $q \geq 2$ by:*

$$\begin{aligned} Bias\left(\bar{f}_b^{G_1}(x)\right) &= \frac{1}{2} x f_0(x) \left[r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\right] b + o(b), \\ Bias\left(\bar{f}_b^{G_2}(x)\right) &= \frac{1}{2} x f_0(x) \left[r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\right] b + o(b), \\ Bias\left(\bar{f}_b^{IG}(x)\right) &= \frac{1}{2} x^3 f_0(x) \left[r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\right] b + o(b), \\ Bias\left(\bar{f}_b^{RIG}(x)\right) &= \frac{1}{2} x f_0(x) \left[r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\right] b + o(b). \end{aligned}$$

The variances of the asymmetric LMBC kernel estimator are the same as those in the pure nonparametric case.

4.3 Special cases

After developing the general LMBC framework, properties of the special cases as described in H&G, H&J and Loader (1996) are now derived. As described in Section 2, direct local modelling of the density can be obtained by choosing the parametric start density as an "improper" uniform distribution. W.l.o.g. we can set $f_0(x)$ to one. The local model $r(t, \theta)$ is then the only source of bias reduction. As soon as we fit more than two local parameters, the bias of the asymmetric local likelihood estimator is $\frac{1}{2}x^a \left[r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\right] b$, where a is equal to three for the inverse Gaussian and one for the other asymmetric kernels. The asymmetric version of the multiplicatively corrected kernel estimator of H&G emerges from choosing the weighting function as $1/f\left(t, \hat{\theta}_1\right)$ and choosing the local model as a constant. From Equation (4.2) it follows that the estimator is in this case

$$\tilde{f}_A(x) = f(x, \hat{\theta}_1)\hat{r}(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(x, b)(X_i) \frac{f(x, \hat{\theta}_1)}{f(X_i, \hat{\theta}_1)}.$$

This estimator has the advantage that it is very simple to implement. Chen's asymmetric kernel estimator has therefore an implicit initial parametric start which is given by an ”improper” uniform distribution. The ratio $f(x, \hat{\theta})/f(X_i, \hat{\theta})$ equals one in this case. This time, no boundary correction terms are needed which contrasts with symmetric kernels. This is because the asymmetric kernel already answers the boundary bias issue. A similar multiplicative bias correction approach developed by Jones, Linton and Nielsen (1995) (JLN) could be extended to the asymmetric case[4]. The analogue of their estimator for asymmetric kernels is

$$\bar{f}_A(x) = \hat{f}_b(x)\hat{\alpha}(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(x, b)(X_i) \frac{\hat{f}_b(x)}{\hat{f}_b(X_i)},$$

where $\hat{f}_b(x)$ is the usual asymmetric kernel estimator given in Equation (3.1). For symmetric kernels, this estimator has generally smaller bias than the standard kernel estimator at the cost of a slightly larger variance. In a simulation study for symmetric kernels Jones and Signorini (1997) find that this method is the overall best among a range of different bias reduction techniques. It did not, however, perform well for the highly skewed test density #3 of Marron and Wand (1992), which is close to lognormal and therefore close to what loss and income distributions are expected to be. Considering this, it is therefore no surprise that Bolancé, Guillen and Nielsen (2003) find that their transformation approach is more suitable for nonparametric loss distribution estimation than the JLN-estimator. They find that this estimator, choosing a constant bandwidth for the whole support, is undersmoothed in

[4] This estimator can be obtained by choosing a nonparametric start estimator and choosing $1/\hat{f}_b(x)$ instead of $1/f_b\left(x, \hat{\theta}_1\right)$ as the weighting function. However, this time the first estimation step does influence the variance of the final estimator, and we can therefore not embed the JLN in the LMBC framework.

the tails of the density. We expect the performance of the asymmetric JLN-estimator to be much better because of the flexible kernel shape. We do, however, not further pursue this idea here since the JLN bias correction procedure is fully nonparametric. It should be possible to derive properties of this estimator in the asymmetric case combining techniques given in JLN and Chen (2000).

4.4 Examples

After developing the general framework, we now show how this framework can be applied to the estimation of densities with support on the nonnegative real line. We focus especially on examples which are relevant for the estimation of loss and income distributions. We give examples for the asymmetric LMBC, the asymmetric version of the H&G estimator and explore also some examples where the local model is chosen to fit the density directly as in H&J and Loader. In Section 5 we compare the finite sample performance of most of these estimators.

4.4.1 A gamma start

A parametric start which is sufficiently flexible and can be expected to be appropriate in applications for unimodal and right skewed distributions is given by the gamma density[5]. This parametric start can be combined with a local polynomial model for the correction factor: $r(t, \theta_2) = \theta_{21} + \theta_{22}(t-x) + ... + \theta_{2(q+1)}(t-x)^q$. The estimator is $\tilde{f}_A(x) = f\left(x, \hat{\theta}_1(x)\right)\hat{\theta}_{21}(x)$. The gamma start density in combination with the first gamma kernel yields easy to implement estimators, which is a further advantage of our method and obviously of considerable importance for practical empirical investigations. Bolancé et. al. (2003) mention that nonparametric methods for loss distribution estimation are seldom applied in practice. Candidate estimators must be easy to implement to have a chance of being applied in the non-academic world.

Example 1 Using the above model for the correction factor, choosing the weight functions $(t-x)^j$ for $j = 0, ..., q$ and using Equation (4.2), one can easily establish that the semiparametric density estimator for a general order q is

$$\tilde{f}(x, q) = f_G\left(x, \hat{\alpha}, \hat{\beta}\right) e_1^t \begin{bmatrix} \delta_0 & ... & \delta_q \\ ... & ... & \ddots \\ \delta_q & ... & \delta_{2q} \end{bmatrix}^{-1} \begin{bmatrix} \hat{f}_b^0(x) \\ ... \\ \hat{f}_b^q(x) \end{bmatrix},$$

[5] An example based on the gamma (and also log-normal) density using symmetric kernels can be found in H&G. Their estimator suffers, however, from boundary bias like standard kernel estimators.

where

$$\delta_j = \int_0^\infty \mathbf{G}\left(x/b+\hat{\alpha}, b\frac{\hat{\beta}}{\hat{\beta}+b}\right)(t-x)^j(t),$$

$$c = \frac{\Gamma\left(\frac{x}{b}+\hat{\alpha}\right)\left(b\frac{\hat{\beta}}{\hat{\beta}+b}\right)^{x/b+\hat{\alpha}}}{\Gamma(\hat{\alpha})\,\hat{\beta}^{\hat{\alpha}}\,\Gamma(x/b+1)\,b^{x/b+1}}, \qquad (4.7)$$

and $\hat{f}_b^j(x)$ is the sample average of $(X_i - x)^j K_{G_1}(x,b)$. Choosing the local model as a constant is not particularly attractive since the bias of this estimator contains also the first order derivative of the correction function and the local model. These first derivative terms vanish if we choose a local polynomial model for the correction factor with $q \geq 2$. In particular, the local linear version of this estimator has the same bias as the asymmetric H&G estimator if K_A is chosen as the RIG, IG or second gamma kernel:

$$\tilde{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\left(\frac{x}{X_i}\right)^{\hat{\alpha}-1}\exp\left\{-\hat{\theta}(x-X_i)\right\}. \qquad (4.8)$$

Both estimators are attractive when the true density is close to the gamma family. Otherwise a local model for the correction factor is desirable which is able to capture its curvature to further diminish the bias. This can be attained by choosing $q \geq 3$. We will consider the first gamma kernel version of the estimator in Equation (4.8) in our simulation study and will refer to it as the AHGG estimator.

Example 2 An alternative to local polynomial modelling of the correction factor is fitting a polynomial to the logarithm of the correction factor, choosing $r(t,\theta) = \theta_1 \exp\left(\theta_2(t-x)+...+\theta_{2(q+1)}(t-x)^q\right)$. Compared to direct polynomial fitting as described above, this ensures a positive estimator and promises a better performance than the H&G estimator if the true density is not given by the parametric start[6]. We work out below the local log linear version of this estimator with a gamma start, again using the first gamma kernel. Using Equation (4.2) and the score as the weighting function, the equation system to be solved is

$$\hat{f}_b(x) = c\theta_1 \exp(-\theta_2 x)\,\psi(\theta_2), \qquad (4.9)$$

$$\hat{f}_b^1(x) = c\theta_1 \exp(-\theta_2 x)\left[\psi^{(1)}(\theta_2) - x\psi(\theta_2)\right], \qquad (4.10)$$

where $\psi(\theta_2)$ is the moment generating function of $\mathbf{G}\left(x/b+\hat{\alpha}, b\frac{\hat{\beta}}{\hat{\beta}+b}\right)$ and c is as defined in (4.7). Since $\tilde{f}(x) = f_G\left(x,\hat{\alpha},\hat{\beta}\right)\hat{\theta}_1,\ \hat{\theta}_2$ is only somewhat "silently" present in the local parameterization. For the first gamma kernel

$$\psi(\theta_2) = (1-\beta^* c)^{-\alpha^*} \qquad \text{for } \theta_2 \leq \beta^{*-1},$$

[6]Note from (4.4) that the leading term of the bias of this estimator is given by $\frac{1}{2}f_0(x)\{r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\}xb$ compared to the H&G estimator which has $\frac{1}{2}f_0(x)\,r_0^{(2)}(x)$ as the leading term.

where $\alpha^* = x/b + \hat{\alpha}$ and $\beta^* = b\frac{\hat{\beta}}{\hat{\beta}+b}$. Using this together with Equations (4.9) and (4.10) yields

$$\hat{\theta}_2 = \frac{(q+x) - \alpha^*\beta^*}{\beta^*(q+x)}, \tag{4.11}$$

where $q = \hat{f}_b^1(x)/\hat{f}_b(x)$[7]. From (4.9), we can then obtain a closed form expression for the LMBC estimator with a gamma start and the log linear correction factor

$$\tilde{f}(x) = f_G\left(x, \hat{\alpha}, \hat{\beta}\right) \frac{\hat{f}_b(x)}{c\exp\left(-\hat{\theta}_2 x\right)\psi\left(\hat{\theta}_2\right)}. \tag{4.12}$$

So with a gamma start this estimator is clearly simple to implement and should reveal appropriate in many circumstances because of the shape flexibility of the gamma distribution. We will refer in the Monte Carlo Section to the estimator of Equation (4.12) as the ALMBC estimator. Unfortunately this simplicity does not extend to other parametric starts than the gamma density. There the integrals corresponding to those in Equation (4.2) have to be calculated numerically.

Alternatively, we also derive an estimator which fits locally a gamma density. This could be of interest when the parameters of the density have some economic interpretation, as in the case of measures of inequality for example.

Example 3 Let us fit a running $\mathbf{G}(\alpha, \beta)$ using 1 and $(t-x)$ as weight functions. This procedure is expected to be appropriate for many right skewed unimodal densities and the estimation procedure proves to be tractable. From (4.2) the equations to solve for $\hat{\alpha}(x)$ and $\hat{\beta}(x)$ are

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\begin{pmatrix} 1 \\ X_i - x \end{pmatrix} \\ = &\int_0^\infty K_A(x,b)(t)\begin{pmatrix} 1 \\ t-x \end{pmatrix}\frac{1}{\Gamma(\alpha)\beta^\alpha}t^{\alpha-1}\exp(-t/\beta)\,dt. \end{aligned} \tag{4.13}$$

Choosing the the first gamma kernel as K_A, we establish in the appendix that this amounts to solving

$$\hat{f}_b(x) = \left(\frac{b}{\beta}\right)^{\alpha(\beta)} \frac{1}{B(\alpha(\beta), x/b)\,x}\left(\frac{\beta}{\beta+b}\right)^{x/b+\alpha(\beta)} \tag{4.14}$$

numerically for β, where $\alpha(\beta) = (b\beta)^{-1} q(\beta+b) + xb$ and $B(\gamma,\delta)$ denotes the Beta function.

One may alternatively use the local likelihood function (4.1) to yield estimators $\hat{\alpha}(x)$ and $\hat{\beta}(x)$. This amounts to maximizing

$$-\log\Gamma(\alpha) - \alpha\log\beta + \frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\left[(\alpha-1)\log X_i - \frac{X_i}{\beta}\right] - d(\theta)$$

[7]It can easily be checked that the solution $\hat{\theta}_2$ always satisfies the restriction to be smaller than $1/\beta^*$.

with $d(\theta) = \hat{f}_b(x)$ as above. One could use for example Newton-Raphson to solve the two equations that use the score of the local likelihood function as the weight functions. These two equations are

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\ln X_i - \hat{f}_b(x)(\text{Psi}(\alpha) + \ln\beta) \\ = \; & d(\theta)\cdot\left[\text{Psi}(x/b+\alpha) - \text{Psi}(\alpha) - \ln\beta + \ln\left(b\frac{\beta}{\beta+b}\right)\right], \end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\frac{1}{\beta}\left(\frac{X_i}{\beta} - \alpha\right) = \frac{d(\theta)}{\beta}\left(\frac{x+b\alpha}{\beta+b} - \alpha\right),$$

where $Psi(\alpha)$ denotes the psi function. For both procedures the running parameter estimates must be computed over a grid of x values. From a practical point of view it may be useful to start equation solving at a new x at the optimized values for the previous x.

4.4.2 Lognormal, Weibull and other start densities

Whereas the integral in Equation (4.2) can be analytically evaluated when we use a gamma kernel in combination with a gamma start, this is no longer true for other popular densities which have support on the nonnegative real line. There, numerical integration techniques are required. It is therefore convenient to choose the H&G weight function which automatically solves this problem. This simplicity of the H&G estimator makes this approach particularly attractive. We develop here a few examples based on parametric start densities which are used in the literature for income and loss distribution modelling. We follow the notation of Klugman et al. (1998).

Example 4 One popular parametric model for loss and income distributions is given by the lognormal probability law $LN(\mu,\sigma)$. This parametric model has the best overall form to fit loss data. The estimator is

$$\hat{f}_A(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\frac{\exp\left\{-\frac{1}{2}(\log x - \hat{\mu})^2/\hat{\sigma}^2\right\}}{\exp\left\{-\frac{1}{2}(\log X_i - \hat{\mu})^2/\hat{\sigma}^2\right\}}\frac{X_i}{x}.$$

Example 5 Another useful parametric start is the Weibull $W(\theta,\tau)$ probability law. The exponential density results if $\tau = 1$.

$$\hat{f}_A(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\left(\frac{X_i}{x}\right)^{1-\hat{\tau}}\exp\left(\left(\frac{1}{\hat{\theta}}\right)^{\tau}\left(X_i^{\hat{\tau}} - x^{\hat{\tau}}\right)\right).$$

Example 6 Finally one can consider a mixture of the above or other distributions as an approximate model for loss data. Hewitt and Lefkowitz (1979) find, for example, that a mixture of the gamma $\mathbf{G}(\alpha,\theta)$

and loggamma(γ,δ) distributions forms an important model for claim distributions. The semiparametric estimator using this implicit start is

$$\hat{f}_A(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\frac{g(x,\hat{\pi})}{g(X_i,\hat{\pi})},$$

where

$$g(x,\hat{\theta}) = \begin{cases} \hat{\psi}\frac{1}{\Gamma(\alpha)\theta^{\alpha}}x^{\alpha-1}\exp(-x/\theta) & \textit{if } 0 < x \le 1, \\ \hat{\psi}\frac{x^{\alpha-1}\exp(-x/\theta)}{\Gamma(\alpha)\theta^{\alpha}} + \left(1-\hat{\psi}\right)\frac{(\ln x)^{\gamma-1}\delta^{\gamma}}{\Gamma(\gamma)x^{\delta+1}} & \textit{if } 1 < x, \end{cases},$$

with $\hat{\pi} = \left(\hat{\alpha},\hat{\theta},\hat{\gamma},\hat{\delta},\hat{\psi}\right)$ and $0 \le \hat{\psi} \le 1$[8].

Klugman et al. (1998) provide many suitable parametric densities to model loss distributions. Any of them can be used as a parametric start. We show later in this paper how suitable parametric start densities can be selected and also propose a semiparametric specification test to determine whether a density belongs to a particular parametric family.

4.4.3 Direct density modelling

Whereas there exists a huge literature concerning parametric estimation of income and loss distributions, in other areas of research it is often not clear what an appropriate parametric start for the density of interest would be. In this case, direct local modelling of the density is an alternative option. Loader (1996) concentrates on local polynomial fitting to the logarithm of the density under consideration and H&J argue that this is more attractive in semiparametric terms than direct local polynomial fitting. This is true for densities the log of which take polynomial form like the normal and exponential density. It is, however, less obvious why one approach should provide a better local fit than the other for densities such as the gamma, Pareto and lognormal whose log is not polynomial. We develop here examples of both versions.

Example 7 A natural idea is local polynomial fitting: $f(t,\theta) = \theta_1 + \theta_2(t-x) + ... + \theta_{q+1}(t-x)^q$. If we choose the weight functions as $(t-x)^j$ for $j = 0,...,q$, then

$$\tilde{f}^A(x,p) = \hat{\theta}_1(x) = e_1'\begin{bmatrix} \delta_0 & ... & \delta_q \\ ... & ... & ... \\ \delta_q & ... & \delta_{2q} \end{bmatrix}^{-1}\begin{bmatrix} \hat{f}_b^0(x) \\ ... \\ \hat{f}_b^q(x) \end{bmatrix}, \tag{4.15}$$

where

$$\delta_j = \int_0^{\infty} K_A(x,b)(t)(t-x)^j\, dt.$$

[8] For computational purposes it is useful to define $\psi = \exp(u)/(1+\exp(u))$ to avoid numerical solutions of ψ which are outside $[0,1]$.

Fitting locally a constant produces the standard asymmetric kernel estimator. This remains true for local linear fitting for the IG and RIG kernel density estimator. It is, however, not the case for the first gamma kernel estimator. This is because the gamma kernel is not centred at x like the RIG and IG kernels, but at $x+b$. For the first gamma kernel, we show in the appendix that local linear fitting yields

$$\tilde{f}^{G_1}(x) = \hat{f}_b(x)\left(\frac{x+2b}{x+b}\right) - \frac{\hat{f}_b^1(x)}{x+b}. \tag{4.16}$$

This estimator no longer has the first order derivative term in its bias. Local quadratic fitting can capture the local curvature of the density and therefore successfully reduce the bias of the asymmetric kernel estimator. Using a RIG kernel estimator, burdensome but straightforward calculations in the appendix show that the local quadratic semiparametric density estimator can be defined by the solution to the following linear equation system:

$$\begin{aligned} \hat{f}_b^0(x) &= \theta_1 + \theta_3\left[xb+b^2\right], \\ \hat{f}_b^1(x) &= \theta_2\left(xb+b^2\right) + \theta_3 b^2\left(3x+5b\right), \\ \hat{f}_b^2(x) &= \theta_1\left(b^2+xb\right) + \theta_2\left(3b^2x+5b^3\right) + \theta_3\left(21b^3x+36b^4+3x^2b^2\right). \end{aligned} \tag{4.17}$$

Similar results can be developed for higher order local polynomial fitting.

Example 8 Choose $f(t,\theta) = \theta_1 \exp\left(\theta_2\left(t-x\right)\right)$. This choice of a local log-linear density is attempting to get the right local slope. If we take the score as the weighting function and use the same procedure as in Example 2, the semiparametric density estimator is

$$\tilde{f}(x) = \frac{\hat{f}_b(x)}{\exp\left(-\hat{\theta}_2 x\right)\psi\left(\hat{\theta}_2\right)}, \tag{4.18}$$

where $\psi\left(\theta_2\right)$ is the m.g.f. of a $\mathbf{G}\left(x/b+1,b\right)$ and $\hat{\theta}_2$ is given in (4.11) for $\alpha^* = x/b+1$ and $\beta^* = b$. This estimator is, therefore, straightforward to implement and always nonnegative. In our simulation study, we will refer to this estimator as the ALLL estimator. Unfortunately this simplicity does not extend to higher order approximations. There the integrals corresponding to those in Equation (4.2) have to be calculated numerically.

4.5 Choice of bandwidth

The mean square error optimal smoothing parameter at point x for the asymmetric LMBC estimator using the first gamma kernel is in the interior

$$b_{G_1}^*(x) = \left(\frac{1}{2\sqrt{\pi}}\right)^{2/5} x^{-1}\left(\frac{f(x)}{\left\{f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right\}^2}\right)^{2/5} n^{-2/5}.$$

The optimal smoothing parameter is large if the parametric guess is close to the true model. The optimal mean squared error is

$$MSE^{*}_{G_1}(x) = \frac{5}{4}\left(\frac{f(x)}{2\sqrt{\pi}}\right)^{4/5}\left\{f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right\}^{2/5} n^{-4/5},$$

and does not depend on x. The optimal $MSE^{*}_{G_1}(x)$ in the boundary is of a less desirable order. Chen (2000) shows, however, that the impact on the $MISE$ is asymptotically negligible. Therefore, regarding global properties, the optimal bandwidth and mean integrated squared error for the first gamma kernel are:

$$\begin{aligned} b^{**}_{G_1}(x) &= \left(\frac{\int_0^\infty \frac{1}{2\sqrt{\pi}} x^{-1/2} f(x)\,dx}{\int_0^\infty x^2\left[f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right]^2 dx}\right)^{2/5} n^{-2/5}, \\ MISE^{**}_{G_1}(x) &= \frac{5}{4}\frac{\left(\int_0^\infty \frac{1}{2\sqrt{\pi}} x^{-1/2} f(x)\,dx\right)^{4/5}}{\left(\int_0^\infty x^2\left[f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right]^2 dx\right)^{-1/5}} n^{-4/5}. \end{aligned}$$

Hence these estimators achieve the optimal rate of convergence for the MISE within the class of non-negative kernel density estimators. Corresponding expressions for the RIG and IG can be derived similarly.

A popular bandwidth selection method for symmetric kernels is unbiased least squares cross validation (LSCV). The idea of this method is to estimate the MISE of the multiplicatively corrected asymmetric kernel estimator and then minimize this expression with respect to the smoothing parameter. A nearly unbiased[9] estimator of $MISE - \int f(x)^2\,dx$ is

$$LSCV(b) = \int_0^\infty \hat{f}_b(x)^2\,dx - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{b,(i)}(X_i), \tag{4.19}$$

where $\hat{f}_{b,(i)}$ is the estimator constructed from the reduced data set that excludes X_i. For the asymmetric H&G estimator, Equation (4.19) can be shown to evaluate as

$$\frac{1}{n^2}\sum_{i,j}^{n}\frac{1}{f(X_i,\hat{\theta})f(X_j,\hat{\theta})}\int_0^\infty f(x,\hat{\theta})^2 K_A(x,b)(X_i)\,K_A(x,b)(X_j)\,dx$$
$$-\frac{2}{n(n-1)}\sum_i\sum_{j\neq i} K_A(X_i,b)(X_j)\,\frac{f(X_i,\hat{\theta}_{(i)})}{f(X_j,\hat{\theta}_{(i)})},$$

where $\hat{\theta}_{(i)}$ is computed without X_i. One could also consider a varying smoothing parameter. We do not pursue this idea here since asymmetric kernels already vary the amount of smoothing through their

[9] H&G show that in the symmetric case this estimator is nearly unbiased already for small samples. Since the arguments are not different in our case, we refer to their paper.

changing shape. Furthermore, second generation bandwidth selection methods such as the smoothed bootstrap for symmetric kernels could be extended to the asymmetric kernel case. For a survey, see Jones, Marron & Sheather (1996).

4.6 Model diagnostics

The estimated correction factor delivers useful information for model diagnostics. The correction factor should equal one if the parametric start density coincides with the true density. We restrict our analysis in this subsection to the H&G estimator. This is because this estimator is already unbiased under true model conditions. Also, the specification test we propose below based on a parametric bootstrap procedure requires fast computation of the estimator.

H&G propose to check model adequacy by looking at a plot of the correction factor for various potential models with pointwise confidence bands to see if $r(x) = 1$ is reasonable. This plot allows to spot easily where misspecification is locally the largest. For the first gamma kernel estimator the bias and variance of the correction factor are

$$E(\hat{r}(x)) = r(x) + b\left[r^{(1)}(x) + \frac{1}{2}xr^{(2)}(x)\right] + o(b), \tag{4.20}$$

$$Var(\hat{r}(x)) = \begin{cases} \frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}x^{-1/2}\frac{r(x)}{f_0(x)} & \text{if } x/b \to \infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}n^{-1}b^{-1}\frac{r(x)}{f_0(x)} & \text{if } x/b \to \kappa. \end{cases} \tag{4.21}$$

Another possibility, in the symmetric case also proposed by H&G, is to plot the log correction factor $\log\hat{r}(x)$ to see how far it is from zero. We derive bias and variance of this curve in the appendix. From the results given there, a simple graphical goodness-of-fit emerges: plot x against

$$Z(x) = \begin{cases} \dfrac{\log\hat{r}(x) + \left(4\sqrt{\pi}n\right)^{-1}(bx)^{-1/2}f\left(x,\hat{\theta}\right)^{-1}}{\left\{\left(2\sqrt{\pi}n\right)^{-1}(bx)^{-1/2}f\left(x,\hat{\theta}\right)^{-1}\right\}^{1/2}} & \text{if } x/b \to \infty, \\ \dfrac{\log\hat{r}(x) + \Gamma(2\kappa+1)\left(2^{2(1+\kappa)}\Gamma^2(\kappa+1)nb\right)^{-1}f\left(x,\hat{\theta}\right)^{-1}}{\left\{\Gamma(2\kappa+1)(2^{1+2\kappa}\Gamma^2(\kappa+1)nb)^{-1}f\left(x,\hat{\theta}\right)^{-1}\right\}^{1/2}} & \text{if } x/b \to \kappa. \end{cases} \tag{4.22}$$

When the parametric start coincides with the true density, this is approximately distributed as standard normal for each x, meaning that the curve should move within ± 1.96 about 95% of the time.

Figure 2 provides an example. 500 random values from a Gamma $(1.5, 1)$ were drawn and the density was estimated by the asymmetric HG estimator with a gamma and a Weibull start. Figure 2 shows that both estimators perform well. As expected, the correction factor for the gamma start estimator is close to one, whereas some nonparametric correction is done for the Weibull start estimator, especially in the tail of the density. Figure 3 plots for both estimators the Z-statistic given in (4.22). Whereas the Z-statistic for the gamma start estimator is always within the confidence bands, the Weibull start estimator is outside at some of the points. The violation is not large. This is because the Weibull can capture the above gamma specification fairly well. Figure 4 shows the same procedure when the true density is LN$(0, 1)$. The correction factors for both estimators indicate that neither a gamma nor a Weibull can capture the tail of lognormal data. Also the close fit shows that although the parametric start is clearly wrong, the density is fitted quite well due to the nonparametric correction for misspecification.

The present framework can also be used to test if the data was generated by a particular parametric model $f(x, \theta)$ where $\theta \in \Theta$ (H_0). We use the global test statistic

$$T_n = nb^{1/2} \int_0^\infty \varphi(x) [\hat{r}(x) - 1]^2 \, dx, \tag{4.23}$$

where $\varphi(x)$ is some appropriately chosen weighting function. Asymptotic normality of this statistic could be shown using results given in Fernandes and Monteiro (2003). It is, however, well known that similar tests based on symmetric kernel estimators are very sensitive to the choice of the smoothing parameter. Fan (1995, 1998) reports that for a wide range of values of the smoothing parameter the test statistics can have large skewness and kurtosis exhibiting behaviour more like χ^2 tests than normal tests. Size distortions can therefore be quite large. He shows that the parametric bootstrap can solve these problems, and we therefore propose the following standard procedure to determine the critical value of the test:

Step 1. Draw a random sample of size n, $\{X_j^*\}_{j=1}^n$, from the distribution with density function $f\left(x, \hat{\theta}\right)$ where $\hat{\theta}$ is estimated by maximum likelihood from the original data. This is the bootstrap sample. Hence conditional on the random sample $\{X_j\}_{j=1}^n$, the boostrap sample satisfies H_0 with $\theta = \theta_0$.

Step 2. Use the bootstrap sample $\{X_j^*\}_{j=1}^n$ in place of the original data to compute T_n. Call it T_n^*.

Step 3. Repeat Step 1 and Step 2 for a large number of times, say B, and obtain the empirical distribution function of $\{T_{n,r}^*\}_{r=1}^B$, called the bootstrap distribution.

Let C_α be the upper $\alpha-$percentile of the calculated bootstrap distribution. Then reject the null hypothesis at significance level α if $T_n > C_\alpha$. T_n is small under the null hypothesis for two reasons. First because the parametric model is correct and $\hat{r}(x)$ should be close to one. Second because the estimator is unbiased and should, therefore, be more precisely measured under the null than under the alternative hypothesis. We therefore expect the power of this semiparametric test to be greater than that of pure nonparametric versions of this kind of specification tests.

Furthermore the statistic given in (4.23) can be used for an adequate choice of a parametric start. The density under consideration can be estimated with different parametric starts. Then one can choose that parametric start density for which the value of the above statistic is the smallest.

4.7 Extensions

Before turning to the Monte Carlo results, we finally would like to mention that our approach can easily be extended to the estimation of densities which have support on the interval $[0,1]$. An application in credit risk is the estimation of the recovery rate density at default, see Renault and Scaillet (2003). To accommodate two known boundaries, Chen (1999) introduced asymmetric kernels based on the beta distribution $\mathbf{B}(x/b+1,(1-x)/b+1)$ given by

$$K_{(x/b+1,(1-x)/b+1)}(t) = \frac{t^{x/b}(1-t)^{(1-x)/b} I(0 \leq t \leq 1)}{B\{x/b+1,(1-x)/b+1\}}.$$

The support of the kernel again matches the support of the density and the resulting estimates are free of boundary bias. An obvious parametric start is given by the beta family of densities. Writing Equation (4.2) using a beta kernel and performing analogous calculations as before, one can establish that the bias and variance of the resulting estimator for $q \geq 2$ are

$$\begin{aligned} Bias\left(\tilde{f}_B(x)\right) &= \frac{1}{2}x(1-x) f_0(x)\left[r^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right] b + o(b), \\ Var\left(\tilde{f}_B(x)\right) &= \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} \{x(1-x)\}^{-1/2} f(x) & \text{if } x/b \to \infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)} n^{-1} b^{-1} f(x) & \text{if } x/b \to \kappa. \end{cases} \end{aligned}$$

The variance of this semiparametric density estimator coincides with that of the pure nonparametric beta kernel estimator. Compared to the bias of the nonparametric beta kernel estimator, first order derivative terms of the true density vanish (as $q \geq 2$) in the bias expression of the semiparametric estimator. Also, $f^{(2)}(x)$ is replaced by $f_0(x)\left[r^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]$. The same remarks apply for the comparison of these biases as before.

Finally the LMBC approach could be extended to a multivariate setting. Since no results have been developed in this setting for standard nonparametric asymmetric kernel estimators, it is beyond the scope of this paper to discuss semiparametric counterparts. We leave this to future research.

5 Monte Carlo results

In this section we evaluate the finite sample performance of some of the estimators considered in the previous section. To the best of our knowledge, it is the first time that various semiparametric density estimators are compared on a finite sample basis.

We run a Monte Carlo simulation for

- the pure nonparametric first gamma kernel estimator (G1),

- the uncorrected semiparametric HG estimator with a gamma start, using the Epanechnikov kernel (SHGG),
- the local linear HG estimator with a gamma start given in Equation (2.6), using the Epanechnikov kernel (SHGGC),
- the semiparametric HG estimator with a gamma start given in Equation (4.8), using the first gamma kernel (AHGG),
- the LMBC estimator with a gamma start and a log linear correction factor, given in Equation (4.12) (ALMBC),
- the local log linear estimator using the first gamma kernel given in Equation (4.18) (ALLL).

We compare these estimators on three different test densities: a Gamma $\mathbf{G}(1.5, 1)$, a Weibull $W(1, 1.5)$ and a lognormal $LN(0, 1)$. The G1 estimator takes the role of the benchmark for the other estimators. A useful semiparametric estimator should at least in some cases outperform its pure non-parametric competitor. Apart from the ALLL estimator, all semiparametric estimators considered here use a parametric gamma start. All should therefore perform well for the gamma test density. We also expect their performance to be good for the Weibull test density, since the gamma start can come close to a Weibull density. The plots of the pseudo gamma and Weibull densities are shown in the left panel of Figure 5[10]. SHGGC is a direct competitor to AHGG. Indeed they are both free of boundary bias. We will see that the HG estimator with a symmetric kernel without boundary correction (SHGG) yields in fact very unsatisfactory results.

In a first simulation step the SHGGC estimator surprisingly performed much worse than the uncorrected estimator SHGG. The reason was that the correction factor can sometimes become too influential. Following similar ideas like H&G, we implemented a trimmed version of the estimator given in Equation (2.6) using

$$\begin{aligned} f\left(x, \hat{\theta}_1\right) \hat{r}(x) &= \frac{1}{nh} \sum_{i=1}^{n} K_h(X_i - x) \min\left(\frac{f\left(x, \hat{\theta}_1\right)}{f\left(X_i, \hat{\theta}_1\right)}, 10\right), \\ f\left(x, \hat{\theta}_1\right) \hat{g}(x) &= \frac{1}{nh} \sum_{i=1}^{n} K_h(X_i - x)(X_i - x) \min\left(\frac{f\left(x, \hat{\theta}_1\right)}{f\left(X_i, \hat{\theta}_1\right)}, 10\right). \end{aligned}$$

[10] The pseudo gamma parameter values are calculated via Monte Carlo integration based on a sample of one million Weibull or lognormal random values.

This trimming procedure successfully solved the numerical problems for this symmetric kernel based estimator. The semiparametric asymmetric kernel estimators did not suffer from the same problem and were implemented without any trimming in the form described in Section 4.

The ALMBC estimator is chosen to examine its performance when the gamma start is clearly wrong, as this is the case for the lognormal test density. This is shown in the right panel of Figure 5, where the pseudo gamma and the true lognormal density are plotted. Also recall our example in Figure 4. The additional local model for the correction factor should theoretically lead to an improvement over the AHGG estimator. Finally, the performance of the ALLL estimator is of interest because it does not need any parametric start but provides a pure local bias correction.

For comparison purposes, we also tried to implement a local log linear estimator with a symmetric kernel. However, this estimator was not suitable for a large scale simulation study, since computation in the boundary of the density requires numerical search procedures in each single step. This is also the reason, why we do not consider the local gamma estimator of Example 3.

The performance measures we consider are the integrated squared error (ISE) and the weighted integrated squared error (WISE) of the various estimators:

$$\begin{aligned} ISE &= \int_{-\infty}^{+\infty} \left\{\bar{f}(x) - f(x)\right\}^2 dx, \\ WISE &= \int_{-\infty}^{+\infty} \left\{\bar{f}(x) - f(x)\right\}^2 x^2 dx. \end{aligned}$$

The WISE allows us to capture the tail performance of our estimators. The experiments are based on 1000 random samples of length $n = 200$, $n = 500$ and $n = 1000$. We provide a "best case" analysis, meaning that for each simulated sample the ISE was computed over a grid of bandwidths and the minimum value was chosen. The WISE is computed in each simulation step with the same bandwidth as the ISE[11]. Numerical integration was performed by Gauss Legendre quadrature with 96 knots.

Table 1 shows the simulation results for the MISE criterion. The AHGG estimator brings large improvements over the G1 estimator when the parametric start is true or close to the true density. Its performance is still slightly better for the lognormal example. This is because the uniform distribution, which is the implicit start for the G1 estimator, is quite a conservative start. The performance of the ALMBC is especially interesting when the parametric start is a poor specification for the true density. Whereas the ALMBC shows as expected a similar performance as the AHGG estimator for the gamma and Weibull test densities, the additional local model for the correction factor brings another 20%

[11] Another procedure would be to compute the WISE as well over a grid of bandwidths and choose the minimum value in each simulation step. We do not follow this because we want to evaluate the tail performance of our estimators given that they fit the whole density well. This is achieved by computing the WISE with the ISE-minimizing bandwidth in each simulation step.

improvement for the lognormal test density. The ALMBC estimator also performs uniformly better than the ALLL estimator, which yields compared to the G1 estimator an improvement between 20-25% across all test densities and sample sizes. The ALLL however does not rely on a parametric start and may perform better when misspecification is stronger than the one considered here.

The poor performance of the SHGG estimator demonstrates how important the boundary bias feature is in the semiparametric framework considered in this paper. Even when the parametric start is correct, SHGG performs worse than the pure nonparametric G1 estimator. Partly, this is because the boundary bias prevents an enlarged bandwidth. Although the trimmed local linear version of this estimator brings a large improvement compared to the uncorrected symmetric estimator, its performance is considerably lower than that of the AHGG estimator. In the slightly misspecified Weibull case with a sample size of 1000, the MISE of the AHGG estimator shrinks to 46% of the MISE of the SHGGC estimator. Chen (2000) has already reported that the asymmetric kernel estimator performs better than its symmetric local linear competitor. The outperformance in our case is however much larger, since the boundary bias problem magnifies in our semiparametric framework as mentioned earlier.

Table 2 shows the same information but for the WISE criterion and makes the power of the asymmetric estimators obvious. Those estimators perform much better in the tail of the density than estimators based on symmetric kernels. For the lognormal density which has the largest tail of the considered test densities, the WISE of the AHGG estimator is just one third of the WISE of its symmetric kernel based competitor SHGGC. We expect this relative advantage to increase for densities that have heavier tails than the lognormal density, e.g. Pareto distributions. This tail advantage of the asymmetric kernel is due to its changing shape as one moves away from the boundary. ALMBC exhibits excellent performance also with respect to the WISE criterion.

6 Empirical applications

In this section we illustrate the usefulness of our estimation approach with two empirical applications. The first one deals with health insurance data provided by a large Swiss health insurer. The second application deals with Brazilian income data.

6.1 Application to health insurance data

The Swiss health insurance system is heavily regulated by law. All residents in Switzerland have a compulsory base insurance which covers general health expenses. In the year 2002, approximately one third of total health expenses of 45 billion Swiss Francs were covered by this base insurance, which is offered by different private insurers.

Since the cost structure among different cantons in Switzerland is very different, we focus here on claims generated by residents of the canton of Zurich. The data considered is the net payment per client in the year 2002, covering claims for the base insurance only. We show how our approach can be used to compare the shape of the loss distribution for different subpopulations to better assess each underlying risk.

Table 3 shows the descriptive statistics of our dataset. It is evident that this dataset is highly skewed and exhibits large kurtosis. Also, the average payment varies significantly with the gender and age of the subpopulations. Obviously, the claim structure is also very different depending whether the client lives in Zurich City or in rural area. We note that the "thought of solidarity" in the Swiss health system implies that clients with an age above 26 years pay all the same premium for their base insurance, independent of age and gender.

Figure 6 shows the loss distribution for the whole sample estimated for comparison purposes by the SHGGC estimator and our ALMBC estimator, both using a gamma start. These two estimators were the best symmetric and asymmetric estimators considered in our Monte Carlo Study. To ensure numerical efficiency, the original dataset was divided by 2'000. Then the resulting density has been back transformed. The bandwidth was calculated using the LSCV procedure described in Section 4.5. We report the bandwidth chosen for the transformed data for all considered estimators in Table 4. We also tried the ALLL estimator, but the resulting density cannot be distinguished by eye from that of the ALMBC estimator and is therefore not plotted. The shape of the loss distribution is very typical, we have a peak for the small claim sizes and then a very long tail. At first glance, the estimates for the symmetric and asymmetric estimator seem quite similar. The correction factor in the right panel of Figure 6 shows, however, that the asymmetric estimator produces a smooth tail, whereas the symmetric estimator features a bumpy behaviour. The picture shows that we have to correct around the mode and also in the tails. The correction factor further left to the picture (not shown) is increasing up to a factor of 10. This is because the gamma start cannot fully capture the heavy tail of the underlying density. The imperfect start and also the bandwidth selected for the ALMBC and ALLL estimator implies that the performance of those estimators is quite similar in terms of precision. Both estimators use a larger bandwidth than the first gamma kernel estimator (not plotted), indicating that they both reduce successfully the bias. This allows us to choose a larger bandwidth compared to the pure nonparametric gamma kernel estimator, which reduces the variance of the estimate. Other parametric starts could be used, e.g. a Pareto distribution to capture better the tail of the density.

Figures 7 and 8 show the loss distributions for different subpopulations, they all seem to be very reasonable and as one would expect a priori. In particular, younger people have smaller claims than older people and are less risky. Our estimators capture very well the heavy tail of clients with an age above 55. Also, the loss distributions for the young people subpopulation does not have a mode, but could be unbounded at zero. This because the majority of their claims are very small.

Although from a social point of view it may be human to charge gender and age independent premiums for health insurance, it is hard to understand, why the Swiss system does not allow the charging of location dependent premiums inside cantons. Figure 7 shows that clients living in Zurich City have a completely different risk structure than people living in the close rural neighbourhood. However, premiums within Canton of Zurich are legally restricted to be the same. Of course, to investigate that point more closely, one would have to condition on the age and gender structure more carefully, but the overall picture would hardly change dramatically.

We conclude that our proposed estimators, which are very simple to apply, seem to be a very useful estimation device for risk managers in insurance companies, and should help to design more differentiated premiums whenever allowed.

6.2 Application to Brazilian income data

Our second application concerns the analysis of the income distribution of Brazil in the year 1990. We analyse a large micro data set (n=71'523), which has been collected by the PNAD annual national household survey. The data set is interesting because Brazil is a major world economy (ninth largest GDP) and faces a strong inequality in terms of percentage shares of income accruing to the richest and to the poorest of its population. The evolution of the Brazilian income distribution in the 1980's has been examined by Cowell, Ferreira and Litchfield (1998). The data considered is monthly household income per capita denominated in 1990 cruzeiros. The strong distributional inequality is revealed by the high skewness of the income distribution, we refer to Table 5 for the descriptive statistics.

We start our analysis with the ALBMC estimator, featuring an implicit gamma start. The parameters of the gamma start density, evaluated by maximum likelihood, are given by $\left(\hat{\alpha}, \hat{\beta}\right) = (0.89, 58861)$, which would imply that the Brazilian income density is unbounded at zero. Figure 9 shows however that the ALMBC estimator does not confirm this gesture. The correction factor in the right panel approaches zero to diminish this effect. Also the correction factor indicates that the gamma model underestimates the mode of the true density as well as its tail. The ALLL estimator cannot be distinguished by eye from the ALMBC and is therefore not plotted. The original dataset was divided by 10'000 and the resulting density estimate back transformed. Again, the bandwidths for the different estimators were chosen according to LSCV and are reported in Table 6 for the transformed data. It is interesting to see that the bandwidth chosen for the ALMBC and ALLL estimators are much larger than that for the first gamma kernel estimator (not plotted). This is because both semiparametric estimators can successfully reduce the bias, which allows us to increase the bandwidth and therefore reduces the variance of the estimates.

Cowell et. al. (1998) mention that the Brazilian income distribution is well approximated by a lognormal model. The above results indicate that this is not very likely in the boundary since $f(0)$ seems not to be zero. Apart from this, as can be seen in Figure 9, the lognormal start for an HG type estimator seems to be very appropriate. The relatively large bandwidth chosen by the LSCV procedure also confirms that the lognormal start contains valuable information.

At this stage we also provide a formal test for lognormality of the underlying income distribution using the test statistic given in Equation (4.23). The estimated parameters for the lognormal model are given by $(\hat{\mu}, \hat{\sigma}) = (10.20, 1.13)$. We use $\varphi(x) = f\left(x, \hat{\theta}\right)$ as a weighting function. This implies that we impose a heavy penalty if the difference between the semiparametric and parametric density estimate is large at those locations where the parametric model puts a lot of weight. We use the bootstrap procedure described in Section 4.6 with $B = 1000$ to approximate the finite sample distribution of the test statistic, using the empirical bandwidth chosen for this data set. We plot the bootstrap density, estimated by the ALLL estimator, of the test statistic under the null hypothesis of lognormality in Figure 10. Although the sample is very large, this bootstrap density is far from normal. This confirms that it is better to rely on a parametric bootstrap rather than relying on asymptotic results. The sample test statistic is given by 29.96 which compares to a critical value of 4.03 at the 1% level. So we reject the null hypothesis of lognormality, which is not surprising since we work with a large sample size and it is unlikely that the underlying density can be described by just two parameters. However as we demonstrated above, the lognormal start contains very valuable information for our semiparametric modelling.

7 Concluding remarks

In this paper we have presented a semiparametric estimation framework based on asymmetric kernels for the estimation of densities on the nonnegative real line. This framework allows us to use popular parametric models from the field of actuarial science and income distribution estimation in a nonparametric fashion. Although the approach per se looks cumbersome, it reduces in many important cases to estimators that take closed forms and are thus very easy to implement. Our simulation results show that our estimators, especially the ALMBC estimator with a parametric start and a local model for the correction factor, exhibit excellent performance. They should therefore be useful in applied work in economics, finance and actuarial science involving non- and semiparametric techniques. This point has already been demonstrated with two empirical applications to health insurance data and Brazilian income data. The results developed here could also be exploited with straightforward modifications in regression curve and hazard rate estimation.

8 Appendix

8.1 Bias and variance of the LMBC estimator

Along the same lines as H&J, we start with the asymptotics for a fixed smoothing parameter. The estimator $\hat{\theta}_2(x)$ we consider is the solution to Equation (4.2). A Taylor expansion around θ_2^0 yields

$$(nb^{1/2})^{1/2}(\hat{\theta}_2 - \theta_2^0) \simeq -V_n^*(\theta_2^0)^{-1}(nb^{1/2})^{1/2}V_n(\theta_2^0),$$

where $V_n^*(\theta_2^0)$ is the $p \times p$ matrix of partial derivatives of the $v_j(\theta)$ functions. Paralleling arguments in the supplementary section of HJ[12], we can establish asymptotic normality of the local parameter estimator $\hat{\theta}_2(x)$:

$$\left(nb^{1/2}\right)^{1/2}\left(\hat{\theta}_2 - \theta_2^0\right) \to^d N\left(0, J_b^{-1}M_b\left(J_b^{-1}\right)^t\right),$$

where

$$J_b(x) = \int_0^\infty K_A(x,b)(t)\left[v(t,\theta_0)u(t,\theta_0)^t m(t,\theta_0) + V^*(t,\theta_0)\{f(t) - m(t,\theta_0)\}\right]dt, \tag{8.1}$$

$$\begin{aligned} M_b(x) &= b^{1/2}Var_f\left(K_A(x,b)(X_i)\,v(x,X_i,\theta_2^0)\right) \\ &= b^{1/2}\int_0^\infty K_A(x,b)(t)^2\, v(t,\theta_2^0)v(t,\theta_2^0)^t f(t)dt - b^{1/2}\xi_b\xi_b^t, \end{aligned} \tag{8.2}$$

and $\xi_b = \int_0^\infty K_A(x,b)(t)\,v(t,\theta_2^0)f(t)dt$. By the delta method the asymptotic distribution of $\hat{f}_A(x)$ for a fixed smoothing parameter is

$$\left(nb^{1/2}\right)^{1/2}\left(\tilde{f}_A(x) - m(x,\theta_0)\right) \to N\left(0, m(x,\theta_0)^2 u(x,\theta_2^0)^t J_b^{-1}M_b\left(J_b^{-1}\right)^t u(x,\theta_2^0)\right). \tag{8.3}$$

In a next step we evaluate the above expressions when the smoothing parameter $b \to 0$ as $n \to \infty$. We illustrate results for the first gamma kernel. Note that

$$V(x,\theta_2^0) = \int_0^\infty K_A(x,b)(t)\,v(x,t,\theta_0)f_0(x)\{r(t) - r(t,\theta_2)\}dt = E\left[q(\gamma_x,\theta_0)\right],$$

where $q(\gamma_x,\theta_0) = v(\gamma_x,\theta_2^0)f_0(\gamma_x)\{r(\gamma_x) - r(\gamma_x,\theta_2^0)\}$ is a $p \times 1$ vector and γ_x is a $\mathbf{G}(\frac{x}{b}+1,b)$ random variable. Concentrating on component j of the vector q and noting that $\mu_x = E(\gamma_x) = x + b$ and $Var(\gamma_x) = xb + b^2$, we perform a Taylor expansion around μ_x and obtain

$$E\left[q_j(\gamma_x,\theta_0)\right] = q_j(x,\theta_0) + \left[q_j^{(1)}(x,\theta_0) + \frac{1}{2}xq_j^{(2)}(x,\theta_0)\right]b + o(b), \tag{8.4}$$

[12] We have the additional problem that $V_n(\theta_2^0)$ is indeed given here by $V_n\left(\hat{\theta}_1, \theta_2^0\right)$ and, therefore, depends on the first step estimator. A Taylor expansion of $V_n\left(\hat{\theta}_1, \theta_2^0\right)$ around $\theta_1^0$and the fact that $\hat{\theta}_1$ is $\sqrt{n}$ convergent shows however immediately, that this is not an issue.

which equals zero at θ_2^0 and therefore

$$v_{j,0}(x) f_0(x) \left\{r(x) - r(x, \theta_2^0)\right\} = \left[q^{(1)}(x, \theta_0) + \frac{1}{2} x q^{(2)}(x, \theta_0)\right] b + o(b). \tag{8.5}$$

From (8.5) it follows that $r(x) - r(x, \theta_2^0) = O(b)$. This together with the fact that $E\left(\tilde{f}_A(x)\right) = f_0(x) r(x, \theta_2^0) + O\left(\frac{1}{nb^{1/2}}\right)$ can be used to obtain the bias of the LMBC gamma kernel estimator:

$$\begin{aligned} Bias\left(\tilde{f}_{G_1}(x)\right) &= \frac{1}{v_{j,0}(x)} \left\{ \left[q^{(1)}(x, \theta_0) + \frac{1}{2} x q^{(2)}(x, \theta_0)\right] b + o(b) \right\} + O\left(\frac{1}{nb^{1/2}}\right) \\ &= f_0(x) \left[\{r_0^{(1)}(x) - r^{(1)}(x, \theta_2^0)\} + \frac{1}{2} x \{r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\} \right] b \\ &\quad + \left(\frac{v_{j,0}^{(1)}(x)}{v_{j,0}(x)} f_0(x) + f_0^{(1)}(x) \right) x \{r_0^{(1)}(x) - r^{(1)}(x, \theta_2^0)\} b + o(b) + O\left(\frac{1}{nb^{1/2}}\right), \end{aligned}$$

which is the result given in the text.

In this paragraph we derive the variance in the one-parameter case for the first gamma kernel. We have to evaluate the expressions $J_b(x)$ and $M_b(x)$ as b approaches zero. In a first step

$$\begin{aligned} J_b(x) &= \int_0^\infty K_A(x, b)(t)\, v(t, \theta_0) u(t, \theta_0) m(t, \theta_0) dt + O(b) \\ &= E\left[c(\gamma_x, \theta_0)\right], \end{aligned}$$

where $c(\gamma_x, \theta_0) = v(\gamma_x, \theta_0) u(\gamma_x, \theta_0) m(\gamma_x, \theta_0)$ and γ_x is $\mathbf{G}(\frac{x}{b}+1, b)$ gamma random variable . Proceeding as in (8.4)

$$E\left[c(\gamma_x, \theta_0)\right] = c(x, \theta_0) + \left[c^{(1)}(x, \theta_0) + \frac{1}{2} x c^{(2)}(x, \theta_0)\right] b + o(b).$$

In the one parameter case the first integral term in Equation (8.2) is

$$\int_0^\infty K_A(x, b)(t)^2\, v(t, \theta_0)^2 f(t) dt = B_b(x)\, E\left[\eta(\zeta_x, \theta_0)\right],$$

where ζ_x follows a $\mathbf{G}(\frac{2x}{b} + 1, \frac{b}{2})$ random variable, $\eta(\zeta_x, \theta_0) = v(\zeta_x, \theta_0)^2 f(\zeta_x)$ and

$$B_b(x) = \frac{\Gamma(2x/b + 1)/b}{2^{2x/b+1} \Gamma^2(x/b + 1)}.$$

Applying the same trick as in (8.4), one can show that the first term in $M_b(x)$ is $b^{1/2} B_b(x)\, v(x, \theta_0)^2 f(x) + O(b^{3/2})$. Proceeding similarly with the second term in Equation (8.2) we get that

$$\xi_b = w(x, \theta_0) + O(b),$$

where $w(x, \theta_0) = v(x, \theta_0) f(x)$. This yields

$$b^{1/2} \xi_b^2 = b^{1/2} w(x, \theta_0)^2 + O(b^{3/2}).$$

Collecting terms,

$$M_b(x) = b^{1/2}\left[B_b(x)\, v(x,\theta_0)^2 f(x) - w(x,\theta_0)^2\right] + O(b^{3/2}).$$

Having derived these preliminary results and using (8.3), we can now tackle the variance of the asymmetric LMBC gamma kernel estimator in the one parameter case.

$$\begin{aligned} Var\left(\tilde{f}_{G_1}(x)\right) &= \frac{1}{nb^{1/2}}\left[m(x,\theta_0)^2 u(t,\theta_0)^2 J_b^{-2} M_b\right] \\ &= \frac{1}{nb^{1/2}} m(x,\theta_0)^2 u(t,\theta_0)^2 \left(\frac{1}{c(x,\theta_0)+O(b)}\right)^2 \\ &\quad \times \left[b^{1/2}\left[B_b(x)\, v(x,\theta_0)^2 f(x) - w(x,\theta_0)^2\right] + O(b^{3/2})\right] \\ &= \frac{1}{n} B_b(x)\, f(x) - \frac{f(x)^2}{n} + O(b/n). \end{aligned}$$

Using the approximation result for $B_b(x)$ given in Chen (2000) proves Equation (4.5) in the main text.

For the multiple parameter case we follow similar ideas like in H&J. In the symmetric case they assume, as is reasonable, that the weight and score functions are of the form

$$c_1 + c_2 hz + c_3 (hz)^2 + ...$$

Transformed to our setting, this means that they are of the form

$$c_1 + c_2 (t-x) + c_3 (t-x)^2 + ...$$

like this is the case in our examples. Since in the asymmetric case we do not have to reparameterize, we can directly define that our weight and score functions (for small b and close to x) have the canonical form

$$V(t) = \left[1, t-x, (t-x)^2, ..., (t-x)^{p-1}\right]'. \tag{8.6}$$

Letting e_1' denote the $p \times 1$ vector $\begin{bmatrix} 1 & 0... & 0 \end{bmatrix}$, using Equations (8.3) and (8.6) we can write the variance of the LMBC estimator in the multiple parameter case as

$$Var\left(\bar{f}^A(x)\right) = \frac{1}{nb^{1/2}} f(x,\theta_0)^2 e_1' J_b^{-1} M_b \left(J_b^{-1}\right)' e_1,$$

where J_b and M_b are given in equations (8.1) and (8.2). Using the above relationships, we begin to work out

$$\begin{aligned} J_b(x) &= \int_0^\infty K_A(x,b)(t)\, v(t,\theta_0) u(t,\theta_0)' m(t,\theta_0) dt + O(b) \\ &= \int_0^\infty K_A(x,b)(t)\, V(t) V(t)' m(t,\theta_0) dt + O(b). \end{aligned}$$

Then from (8.2)

$$M_b(x) = b^{1/2} \int_0^\infty K_A(x,b)(t)^2\, V(t) V(t)' f(t) dt - b^{1/2} \xi_b \xi_b',$$

and

$$\begin{aligned}\xi_b &= \int_0^\infty K_A(x,b)(t)\, V(t) f(t) dt \\ &= e_1 f(x) + O(b).\end{aligned}$$

So

$$b^{1/2}\xi_b\xi_b^{'} = b^{1/2} e_1 e_1^t f(x)^2 + O\left(b^{3/2}\right).$$

Substituting these results in (8.3) we obtain with a slight abuse of notation that

$$\begin{aligned}Var\left(\bar{f}^A(x)\right) &= \frac{1}{n} m(x,\theta_0)^2 e_1^{'} J_b^{-1} \int_0^\infty K_A(t)^2\, V(t)V(t)^{'} f(t) dt \left(J_b^{-1}\right)^{'} e_1 \\ &\quad - \frac{1}{nb^{1/2}} m(x,\theta_0)^2 e_1^{'} J_b^{-1} \left(b^{1/2} e_1 e_1^{'} f(x)^2 + O\left(b^{3/2}\right)\right) \left(J_b^{-1}\right)^{'} e_1 \\ &= \frac{1}{n} m(x,\theta_0)^2 e_1^{'} J_b^{-1} \int_0^\infty K_A(t)^2\, V(t)V(t)^{'} f(t) dt \left(J_b^{-1}\right)^{'} e_1 - \frac{f(x)^2}{n} + O(b/n). \quad (8.7)\end{aligned}$$

Note that

$$\int_0^\infty K_A(t)\, V(t)V(t)^{'} m(t,\theta_0) dt = m(x,\theta_0) \int_0^\infty K_A(t)\, V(t)V(t)^{'} dt + O(b),$$

and similarly

$$\int_0^\infty K_A(t)^2\, V(t)V(t)^{'} f(t) dt = f(x) \int_0^\infty K_A(t)^2\, V(t)V(t)^{'} dt + O(b).$$

Using this, the first term in Equation (8.7) can be simplified to

$$\frac{f(x)}{n} e_1^{'} \tau(K_A)\, e_1 + O(b/n),$$

where $\tau(K_A)$ is

$$\left(\int_0^\infty K_A(t)\, V_t V_t^{'} dt\right)^{-1} \left(\int_0^\infty K_A(t)^2\, V_t V_t^{'} dt\right) \left(\int_0^\infty K_A(t)\, V_t V_t^{'} dt\right)^{-1}.$$

The variance in the multiple parameter case can therefore be written as in (4.6). In the first gamma kernel case and for two local parameters, we have that $\tau(K_A)$ is

$$B_b(x) \begin{bmatrix} 1 & b \\ b & xb + 2b^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & \frac{1}{2}b \\ \frac{1}{2}b & \frac{1}{2}xb \end{bmatrix} \begin{bmatrix} 1 & b \\ b & xb + 2b^2 \end{bmatrix}^{-1},$$

and the upper left element which is of interest can be evaluated as $B_b(x)\,(1 + O(b))$. In the case of two parameters we therefore get the same variance as in the single parameter case. This can also be shown for the other asymmetric kernel estimators.

8.2 Local Gamma density estimation

We develop here Equation (4.14) in the text. From Equation (4.13) the first integral is

$$c \int_0^\infty t^{x/b+\alpha-1} \exp\left(-t\left(\frac{1}{b}+\frac{1}{\beta}\right)\right) dt,$$

where $c = 1/\left(\Gamma(\alpha)\,\beta^\alpha \Gamma(x/b+1)\, b^{x/b+1}\right)$. The right hand side is part of the $\mathbf{G}\left(x/b+\alpha, b\frac{\beta}{\beta+b}\right)$ probability density function and, therefore, the first equation of the two dimensional nonlinear system is

$$\hat{f}_b(x) = \frac{\Gamma\left(\frac{x}{b}+\alpha\right)\left(b\frac{\beta}{\beta+b}\right)^{x/b+\alpha}}{\Gamma(\alpha)\,\beta^\alpha \Gamma(x/b+1)\, b^{x/b+1}}. \tag{8.8}$$

Note that

$$\frac{\Gamma\left(\frac{x}{b}+\alpha\right)}{\Gamma(\alpha)\,\Gamma(x/b+1)} = \frac{\Gamma\left(\frac{x}{b}+\alpha\right) b}{\Gamma(\alpha)\,\Gamma(x/b)\, x} = \frac{b}{B(\alpha, x/b)\, x},$$

where $B(.,.)$ is the beta function. So (8.8) yields

$$\begin{aligned}\hat{f}_b(x) &= \frac{b}{B(\alpha, x/b)\, x} \frac{\left(b\frac{\beta}{\beta+b}\right)^{x/b+\alpha}}{\beta^\alpha b^{x/b+1}} \\ &= \left(\frac{b}{\beta}\right)^\alpha \frac{1}{B(\alpha, x/b)\, x}\left(\frac{\beta}{\beta+b}\right)^{x/b+\alpha} = s(\theta). \end{aligned} \tag{8.9}$$

The second integral in (4.13) can be evaluated as follows:

$$\begin{aligned} & \frac{1}{\Gamma(\alpha)\,\beta^\alpha} \int_0^\infty K_A(x,b)(t)\,(t-x)\, t^{\alpha-1} \exp(-t/\beta)\, dt \\ = & \frac{1}{\Gamma(\alpha)\,\beta^\alpha} \int_0^\infty t K_A(x,b)(t)\, t^{\alpha-1} \exp(-t/\beta)\, dt - x s(\theta) \\ = & \; c \int_0^\infty t \cdot t^{x/b+\alpha-1} \exp\left(-t\left(\frac{1}{b}+\frac{1}{\beta}\right)\right) dt - x s(\theta), \end{aligned}$$

where the integral is up to a constant part of the expectation for the $\mathbf{G}\left(x/b+\alpha, b\frac{\beta}{\beta+b}\right)$ random variable. This second integral is

$$\frac{\Gamma\left(\frac{x}{b}+\alpha\right)\left(b\frac{\beta}{\beta+b}\right)^{x/b+\alpha}}{\Gamma(\alpha)\,\beta^\alpha \Gamma(x/b+1)\, b^{x/b+1}} \beta\frac{x+b\alpha}{\beta+b} = s(\theta)\,\beta\frac{x+b\alpha}{\beta+b},$$

and it therefore follows that the second equation to solve is

$$\hat{f}_b^1(x) = s(\theta)\left(\beta\frac{x+b\alpha}{\beta+b} - x\right). \tag{8.10}$$

From Equations (8.9), (8.10) and writing $q = \hat{f}_b^1(x)/\hat{f}_b(x)$,

$$q = \beta\frac{x+b\alpha}{\beta+b} - x.$$

We solve for α in terms of β :

$$\alpha = \frac{q(\beta+b)+xb}{b\beta}.$$

Substituting this result in equation (8.9) yields

$$\hat{f}_b(x) = \left(\frac{b}{\beta}\right)^{\alpha(\beta)} \frac{1}{B(\alpha(\beta), x/b)\, x} \left(\frac{\beta}{\beta+b}\right)^{x/b+\alpha(\beta)},$$

which has to be solved for β. This has to be evaluated numerically.

Let us try now to get a tractable solution in the case where we take the score as the weight functions. We first calculate the score:

$$\log f(t,\theta) = -\log\Gamma(\alpha) - \alpha\log\beta + (\alpha-1)\log t - \frac{t}{\beta}.$$

The first element of the score is

$$\frac{d}{d\alpha}\left(-\log\Gamma(\alpha) - \alpha\log\beta + (\alpha-1)\log t - \frac{t}{\beta}\right) = -\operatorname{Psi}(\alpha) - \log\beta + \log t,$$

while the second element is

$$\frac{d}{d\beta}\left(-\log\Gamma(\alpha) - \alpha\log\beta + (\alpha-1)\log t - \frac{t}{\beta}\right) = \frac{1}{\beta}\left(\frac{t}{\beta}-\alpha\right).$$

The equations we have to solve are

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\begin{pmatrix} -\operatorname{Psi}(\alpha) - \ln\beta + \ln X_i \\ \frac{1}{\beta}\left(\frac{X_i}{\beta}-\alpha\right)\end{pmatrix} \qquad (8.11)\\ = &\int_0^\infty K_A(x,b)(t)\begin{pmatrix} -\operatorname{Psi}(\alpha) - \ln\beta + \ln t \\ \frac{1}{\beta}\left(\frac{t}{\beta}-\alpha\right)\end{pmatrix}\frac{1}{\Gamma(\alpha)\beta^\alpha}t^{\alpha-1}\exp(-t/\beta)\,dt. \end{aligned}$$

Consider the first integral (8.11) :

$$-(\operatorname{Psi}(\alpha)+\log\beta)\,s(\theta) + \int_0^\infty K_A(x,b)(t)\log t\frac{1}{\Gamma(\alpha)\beta^\alpha}t^{\alpha-1}\exp(-t/\beta)\,dt,$$

where $s(\theta)$ is as defined in (8.9). It remains to calculate the last integral. Note that

$$\begin{aligned} &\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty K_A(x,b)\log t\cdot t^{\alpha-1}\exp(-t/\beta)\,dt \\ = \; & c\int_0^\infty \log t\cdot t^{x/b+\alpha-1}\exp\left(-t\left(\frac{1}{b}+\frac{1}{\beta}\right)\right)dt. \end{aligned}$$

This is part of the expectation of $\log t$ with respect to a $\mathbf{G}\left(x/b+\alpha, b\frac{\beta}{\beta+b}\right)$ random variable. We can write as

$$s(\theta)\,E(\log t).$$

We must evaluate

$$E(\log t) = \frac{1}{\Gamma(\gamma)\delta^{\gamma}} \int_0^{\infty} \log t \cdot t^{\gamma-1} \exp(-t/\delta)\, dt,$$

where $\gamma = x/b + \alpha > 0$ and $\delta = b\frac{\beta}{\beta+b} > 0$. This integral is

$$\begin{aligned} & \frac{1}{\Gamma(\gamma)\delta^{\gamma}} \delta^{\gamma}\Gamma(\gamma)(\log\delta + \mathrm{Psi}(\gamma)) \\ = \;& \log\delta + \mathrm{Psi}(\gamma). \end{aligned}$$

The first equation in (8.11) equals

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\log X_i - \hat{f}_b(x)(\mathrm{Psi}(\alpha) + \log\beta) \\ = \;& s(\theta)\left[\mathrm{Psi}(x/b+\alpha) - \mathrm{Psi}(\alpha) - \log\beta + \log\left(b\frac{\beta}{\beta+b}\right)\right]. \end{aligned}$$

Using similar calculations as above we can evaluate the second integral in (8.11) as

$$\begin{aligned} & \int_0^{\infty} K_A(x,b)(t)\frac{1}{\beta}\left(\frac{t}{\beta} - \alpha\right)\frac{1}{\Gamma(\alpha)\beta^{\alpha}} t^{\alpha-1}\exp(-t/\beta)\, dt \\ = \;& \frac{1}{\beta^2}\int_0^{\infty} K_A(x,b)(t)\, t\frac{1}{\Gamma(\alpha)\beta^{\alpha}} t^{\alpha-1}\exp(-t/\beta)\, dt - \frac{\alpha}{\beta}s(\theta) \\ = \;& \frac{1}{\beta^2}s(\theta)\cdot\beta\frac{x+b\alpha}{\beta+b} - \frac{\alpha}{\beta}s(\theta) \\ = \;& \frac{s(\theta)}{\beta}\left(\frac{x+b\alpha}{\beta+b} - \alpha\right). \end{aligned}$$

The second equation in the nonlinear system (8.11) is

$$\frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\frac{1}{\beta}\left(\frac{X_i}{\beta} - \alpha\right) = \frac{s(\theta)}{\beta}\left(\frac{x+b\alpha}{\beta+b} - \alpha\right).$$

8.3 Local polynomial examples

We develop Equation (4.16) in the text. Using 1 and $(t-x)$ as weight functions, in the first gamma kernel case we must solve the equations

$$\frac{1}{n}\sum_{i=1}^{n} K_A(x,b)(X_i)\begin{pmatrix} 1 \\ X_i - x \end{pmatrix} = \int_0^{\infty} K_A(x,b)(t)\begin{pmatrix} 1 \\ t - x \end{pmatrix}(\theta_1 + \theta_2(t-x))\, dt.$$

The first equation can be written as

$$\begin{aligned} \hat{f}_b(x) &= \int_0^{\infty} K_A(x,b)(t)(\theta_1 + \theta_2(t-x))\, dt \\ &= \theta_1 + \theta_2 b, \end{aligned} \tag{8.12}$$

whereas the second equation is

$$\begin{aligned}\hat{f}_b^1(x) &= \int_0^\infty K_A(x,b)(t)\left[\theta_1(t-x)+\theta_2(t-x)^2\right]dt \\ &= \theta_1 b+\theta_2\int_0^\infty K_A(x,b)(t)\left(t^2-2tx+x^2\right)dt.\end{aligned}$$

Since for the $\mathbf{G}(\alpha,\beta)$ we have that $E(t^2)=\beta^2\alpha(\alpha+1)$ this yields

$$\begin{aligned}\int_0^\infty K_A(x,b)(t)\,t^2dt &= b^2\left(\frac{x}{b}+1\right)\left(\frac{x}{b}+2\right) \\ &= x^2+3xb+2b^2,\end{aligned}$$

and therefore

$$\begin{aligned}\hat{f}_b^1(x) &= \theta_1 b+\theta_2\left[x^2+3xb+2b^2-2x(x+b)+x^2\right] \\ &= \theta_1 b+\theta_2\left[xb+2b^2\right].\end{aligned} \tag{8.13}$$

Solving the equation system (8.12), (8.13) yields

$$\begin{aligned}\hat{f}_b(x)-\frac{\hat{g}_b(x)}{b} &= \theta_2 b-\theta_2(x+2b) \\ &= -\theta_2(x+b),\end{aligned}$$

which can be solved for

$$\hat{\theta}_2=\frac{\hat{g}_b(x)}{b(x+b)}-\frac{\hat{f}_b(x)}{(x+b)}. \tag{8.14}$$

Using (8.12) and (8.14) we determine

$$\begin{aligned}\hat{\theta}_1 &= \hat{f}_b(x)-b\left[\frac{\hat{g}_b(x)}{b(x+b)}-\frac{\hat{f}_b(x)}{(x+b)}\right] \\ &= \hat{f}_b(x)\left(\frac{x+2b}{x+b}\right)-\frac{\hat{g}_b(x)}{x+b},\end{aligned}$$

which is the required estimator given in Equation (4.16).

We go on with developing the local quadratic fitting example. From Equation (4.15), we have to calculate integrals of the form $\int_0^\infty K_A(x,b)(t)\,t^j dt$ for $j=1,...,4$, for the $RIG(m,d)$. By direct integration:

$$\begin{aligned}E(t) &= \frac{1}{m}+\frac{1}{d}, \\ E\left(t^2\right) &= \frac{3}{dm}+\frac{3}{d^2}+\frac{1}{m^2}, \\ E\left(t^3\right) &= \frac{15}{d^2m}+\frac{1}{m^3}+\frac{15}{d^3}+\frac{6}{dm^2}, \\ E\left(t^4\right) &= \frac{105}{d^3m}+\frac{10}{dm^3}+\frac{45}{d^2m^2}+\frac{1}{m^4}+\frac{105}{d^4}.\end{aligned}$$

We substitute $d = \frac{1}{b}$ and $m = \frac{1}{(x-b)}$ to get after some manipulations

$$\begin{aligned}
E(t) &= x, \\
E\left(t^2\right) &= xb + b^2 + x^2, \\
E\left(t^3\right) &= 6b^2x + 5b^3 + x^3 + 3x^2b, \\
E\left(t^4\right) &= 41b^3x + 36b^4 + 6bx^3 + 21x^2b^2 + x^4.
\end{aligned}$$

These preliminary calculations allow us now to calculate the three required integrals in Equation (4.17). The first integral is

$$\begin{aligned}
&\int_0^\infty K_A(x,b)(t)\left(\theta_1 + \theta_2(t-x) + \theta_3(t-x)^2\right) dt \\
&= \theta_1 + \theta_3 \int_0^\infty K_A(x,b)(t)(t-x)^2\, dt \\
&= \theta_1 + \theta_3 \left[\int_0^\infty K_A(x,b)(t)\left(t^2 - 2tx + x^2\right) dt\right] \\
&= \theta_1 + \theta_3\left[xb + b^2 + x^2 - 2x^2 + x^2\right] \\
&= \theta_1 + \theta_3\left[xb + b^2\right].
\end{aligned}$$

The second integral is

$$\begin{aligned}
&\int_0^\infty K_A(x,b)(t)\left(\theta_1(t-x) + \theta_2(t-x)^2 + \theta_3(t-x)^3\right) dt \\
&= \int_0^\infty K_A(x,b)(t)\left[\theta_2(t-x)^2 + \theta_3(t-x)^3\right] dt \\
&= \int_0^\infty K_A(x,b)(t)\left[\theta_2 t^2 - 2\theta_2 tx + \theta_2 x^2 + \theta_3 t^3 - 3\theta_3 t^2 x + 3\theta_3 t x^2 - \theta_3 x^3\right] dt \\
&= \theta_2\left(xb + b^2 + x^2\right) - 2\theta_2 x^2 + \theta_2 x^2 + \theta_3\left(6b^2x + 5b^3 + x^3 + 3x^2b\right) \\
&\quad -3\theta_3 x\left(xb + b^2 + x^2\right) + 3\theta_3 x^3 - \theta_3 x^3 \\
&= \theta_2 xb + \theta_2 b^2 + 3\theta_3 b^2 x + 5\theta_3 b^3.
\end{aligned}$$

Finally the third integral to compute is

$$
\begin{aligned}
& \int_0^{\infty} K_A(x,b)(t)\left(\theta_1(t-x)^2+\theta_2(t-x)^3+\theta_3(t-x)^4\right)dt \\
= & \int_0^{\infty} K_A(x,b)(t)\left[\begin{array}{c} \theta_1 t^2-2\theta_1 tx+\theta_1 x^2+\theta_2 t^3 \\ -3\theta_2 t^2 x+3\theta_2 tx^2-\theta_2 x^3+\theta_3 t^4-4\theta_3 t^3 x \\ +6\theta_3 t^2 x^2-4\theta_3 tx^3+\theta_3 x^4 \end{array}\right] dt \\
= & \int_0^{\infty} K_A(x,b)(t)\left[\begin{array}{c} \theta_1 t^2-\theta_1 x^2+\theta_2 t^3-3\theta_2 t^2 x+2\theta_2 x^3+ \\ \theta_3 t^4-4\theta_3 t^3 x+6\theta_3 t^2 x^2-3\theta_3 x^4 \end{array}\right] dt \\
= & \theta_1\left(xb+b^2+x^2\right)-\theta_1 x^2+\theta_2\left(6b^2x+5b^3+x^3+3x^2b\right) \\
& -3\theta_2 x\left(xb+b^2+x^2\right)+2\theta_2 x^3+\theta_3\left(41b^3x+36b^4+6bx^3+21x^2b^2+x^4\right) \\
& -4\theta_3 x\left(6b^2x+5b^3+x^3+3x^2b\right)+6\theta_3 x^2\left(xb+b^2+x^2\right)-3\theta_3 x^4 \\
= & 21\theta_3 b^3 x+\theta_1 b^2+5\theta_2 b^3+36\theta_3 b^4+\theta_1 xb+3\theta_2 b^2 x+3\theta_3 x^2 b^2.
\end{aligned}
$$

Collecting terms, the three equations to solve are

$$
\begin{aligned}
\hat{f}_b(x) &= \theta_1+\theta_3\left[xb+b^2\right], \\
\hat{f}_b^1(x) &= \theta_2\left(xb+b^2\right)+\theta_3 b^2(3x+5b), \\
\hat{f}_b^2(x) &= \theta_1\left(b^2+xb\right)+\theta_2\left(3b^2x+5b^3\right)+\theta_3\left(21b^3x+36b^4+3x^2b^2\right),
\end{aligned}
$$

which is the system of Equations (4.17) given in the text.

8.4 Model diagnostics

In this subsection we derive the expression for the test statistics given in Equation (4.22). In the interior using Equations (4.21) and (4.20), we can write

$$
\begin{aligned}
E(\log\hat{r}(x)) &\simeq \log r(x)+\frac{1}{r(x)}E[\hat{r}(x)-r(x)]-\frac{1}{2r(x)^2}E[\hat{r}(x)-r(x)]^2 \\
&= \log r(x)+\frac{1}{r(x)}E[\hat{r}(x)-r(x)]-\frac{1}{2r(x)^2}\left(Var(\hat{r}(x))+Bias(\hat{r}(x))^2\right) \\
&= \log r(x)+b\left[\frac{r^{(1)}(x)}{r(x)}+\frac{1}{2}x\frac{r^{(2)}(x)}{r(x)}\right]-\frac{1}{4\sqrt{\pi}}n^{-1}b^{-1/2}x^{-1/2}\frac{1}{r(x)f_0(x)}+O(b/n).
\end{aligned}
$$

Similarly the variance is

$$
\begin{aligned}
Var\left(\log \hat{r}\left(x\right)\right) &\simeq Var\left(\log r\left(x\right)+\frac{1}{r\left(x\right)}\left(\hat{r}\left(x\right)-r\left(x\right)\right)\right)\\
&= \frac{1}{r\left(x\right)^{2}}Var\left(\hat{r}\left(x\right)\right)\\
&= \frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}x^{-1/2}\left[r\left(x\right)f_0(x)\right]^{-1}+O\left(b/n\right).
\end{aligned}
$$

Assuming that the parametric model is true, the CLT implies that

$$
\begin{aligned}
Z\left(x\right) &= \frac{\log \hat{r}\left(x\right)-E\left(\log \hat{r}\left(x\right)\right)}{SD\left(\log \hat{r}\left(x\right)\right)}\\
&= \frac{\log \hat{r}\left(x\right)+\left(4\sqrt{\pi}n\right)^{-1}\left(bx\right)^{-1/2}f\left(x,\hat{\theta}\right)^{-1}}{\left\{\left(2\sqrt{\pi}n\right)^{-1}\left(bx\right)^{-1/2}f\left(x,\hat{\theta}\right)^{-1}\right\}^{1/2}}
\end{aligned}
$$

is asymptotically normally distributed. The expression for $Z\left(x\right)$ in the boundary can be derived along the same lines.

Acknowledgements Both authors gratefully acknowledge financial support from the Swiss National Science Foundation through the National Centre of Competence in Research: Financial Valuation and Risk Management (NCCR-FINRISK). We also would like to thank R. Abul Naga, N. Arnold, F. Cowell, C. Gouriéroux, O. Linton, A. Monfort, P. Funk, M. Rockinger, R. Stulz, E. von Thadden, B. Ziemba, participants at the 2003 ESRC Econometric Study Group Meeting in Bristol, the 2003 LAMES in Panama City, the FAME Doctoral Workshop in Lausanne, the Swiss Doctoral Workshop in Finance in Gerzensee and the Economics Lunch seminar in Basle for helpful comments. J. Litchfield and H. Beck kindly provided the Brazilian income and health insurance data. Part of this research was done when the first author was visiting the LSE and the second author THEMA and ECARES.

References

[1] AÏT-SAHALIA, Y. (1996a). "Nonparametric Pricing of Interest Rate Derivative Securities", Econometrica, 64, 527-560.

[2] AÏT-SAHALIA, Y. (1996b). "Continuous-Time Models of the Spot Interest Rate", The Review of Financial Studies, 9, 385-426.

[3] BOLANCÉ, C., GUILLEN, M. and J.P. NIELSEN (2003). "Kernel Density Estimation of Actuarial Loss Functions", Insurance: Mathematics and Economics, 32, 3-18.

[4] BOUEZMARNI, T. and O. SCAILLET (2003). "Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data", HEC Genève and Institut de Stat. UCL Discussion Paper 0306.

[5] CHEN, S.X. (1999). "A Beta Kernel Estimator for Density Functions, Computational Statistics and Data Analysis", 31, 131-145.

[6] CHEN, S.X. (2000). "Probability Density Function Estimation Using Gamma Kernels", Annals of the Institute of Statistical Mathematics, 52, 471-480.

[7] COWELL, F. (1999). "Measurement of Inequality", in: Atkinson, A. and F. Bourguignon, eds., Handbook of Income Distribution, North Holland, Amsterdam.

[8] COWELL, F., FERREIRA, F. and J. LITCHFIELD (1998). "Income Distribution in Brazil 1981-1990: Parametric and Non-Parametric Approaches", Journal of Income Distribution, 8, 63-76.

[9] COX, J.C., INGERSOLL, J.E. and S.A. ROSS (1985). "A Theory of the Term Structure of Interest Rates", Econometrica, 53, 385-408.

[10] ENGLE, R.F. (2000). "The Econometrics of Ultra-High Frequency Data", Econometrica, 68, 1-22.

[11] FAN, Y. (1995). "Bootstrapping a Consistent Nonparametric Goodness-of-Fit Test", Econometric Reviews, 14, 367-382.

[12] FAN, Y. (1998). "Goodness-of-Fit Tests based on Kernel Density Estimators with Fixed Smoothing Parameters", Econometric Theory, 14, 604-621.

[13] FERNANDES, M. and J. GRAMMIG (2000). "Nonparametric Specification Tests for Conditional Duration Models", Working Paper, Getulio Vargas Foundation, Rio de Janeiro, Brazil.

[14] FERNANDES, M. and P.K. MONTEIRO (2003). "Central Limit Theorem for Asymmetric Kernel Functionals", Working Paper, Getulio Vargas Foundation, Rio de Janeiro, Brazil.

[15] GOZALO, P. and O.B. LINTON (2000). "Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression", Journal of Econometrics, 99, 63-106.

[16] GOURIEROUX, CH., MONFORT, A. and A. TROGNON (1984). "Pseudo Maximum Likelihood Methods: Theory", Econometrica, 52, 680-700.

[17] GOURIEROUX, CH. and J. JASIAK (2001). "Local Likelihood Density Estimation and Value at Risk", Discussion Paper CREST and CEPREMAP.

[18] HEWITT, CH. and B. LEFKOWITZ (1979). "Methods of Fitting Distributions to Insurance Loss Data", Proceedings of the Casualty Actuarial Society, 66, 139-160.

[19] HJORT, N.L. and I.K. GLAD (1995). "Nonparametric Density Estimation with a Parametric Start", The Annals of Statistics, 24, 882-904.

[20] HJORT, N.L. and M.C. JONES (1996). "Locally Parametric Nonparametric Density Estimation", The Annals of Statistics, 24, 1619-1647.

[21] HOGG, R. and S.A. KLUGMAN (1984). "Loss Distributions", New York: John Wiley & Sons, Inc.

[22] JONES, M.C. (1993). "Simple Boundary Correction for Kernel Density Estimation", Statistics and Computing, 3, 135-146.

[23] JONES, M.C. and P. FOSTER (1996). "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation", Statistica Sinica, 6, 1005-1013.

[24] JONES, M.C., MARRON, J.S. and S.J. SHEATER (1996). "A Brief Survey of Bandwidth Selection for Density Estimation", Journal of the American Statistical Association, 92, 1063-1073.

[25] JONES, M.C., LINTON, O.B. and J.P. NIELSEN (1995). "A Simple but Effective Bias Reduction Method for Density and Regression Estimation", Biometrika, 82, 327-338.

[26] JONES, M.C. and J.P. SIGNORINI (1997). "A Comparison of Higher-Order Bias Kernel Density Estimators", Journal of the American Statistical Association, 92, 1063-1073.

[27] KLUGMAN, S.A., PANJER H.A. and G.E. WILLMOT (1998). "Loss Models: From Data to Decisions", New York: John Wiley & Sons, Inc.

[28] LINTON, O.B. and A. PAKES (2001). "Nonlinear Estimators in Econometrics", manuscript, London School of Economics and Political Science.

[29] LOADER, C.R. (1996). "Local Likelihood Density Estimation", The Annals of Statistics, 24, 1602-1618.

[30] MARRON, J.S. and M.P. WAND (1992). "Exact Mean Integrated Squared Error", The Annals of Statistics, 20, 712-736.

[31] MÜLLER, H. (1991). "Smooth Optimum Kernel Estimators near Endpoints", Biometrika, 78, 521-520.

[32] RENAULT, O. and O. SCAILLET (2003). "On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities", FAME Discussion Paper 83, Switzerland.

[33] RICE, J. (1984). "Boundary Modification for Kernel Regression", Communications in Statistics: Theory and Methods, 13, 893-900.

[34] ROSENBLATT, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function", The Annals of Mathematical Statistics, 27, 642-669.

[35] RUPPERT, D. and M.P. WAND (1994). "Multivariate Weighted Least Squares Regression", The Annals of Statistics, 22, 1346-1370.

[36] SCAILLET, O. (2003). "Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels", forthcoming in Journal of Nonparametric Statistics.

[37] SCHUSTER, E. (1985). "Incorporating Support Constraints into Nonparametric Estimators of Densities", Communications in Statistics: Theory and Methods, 1123-1136.

[38] WHITE, H. (1982). "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1-26.

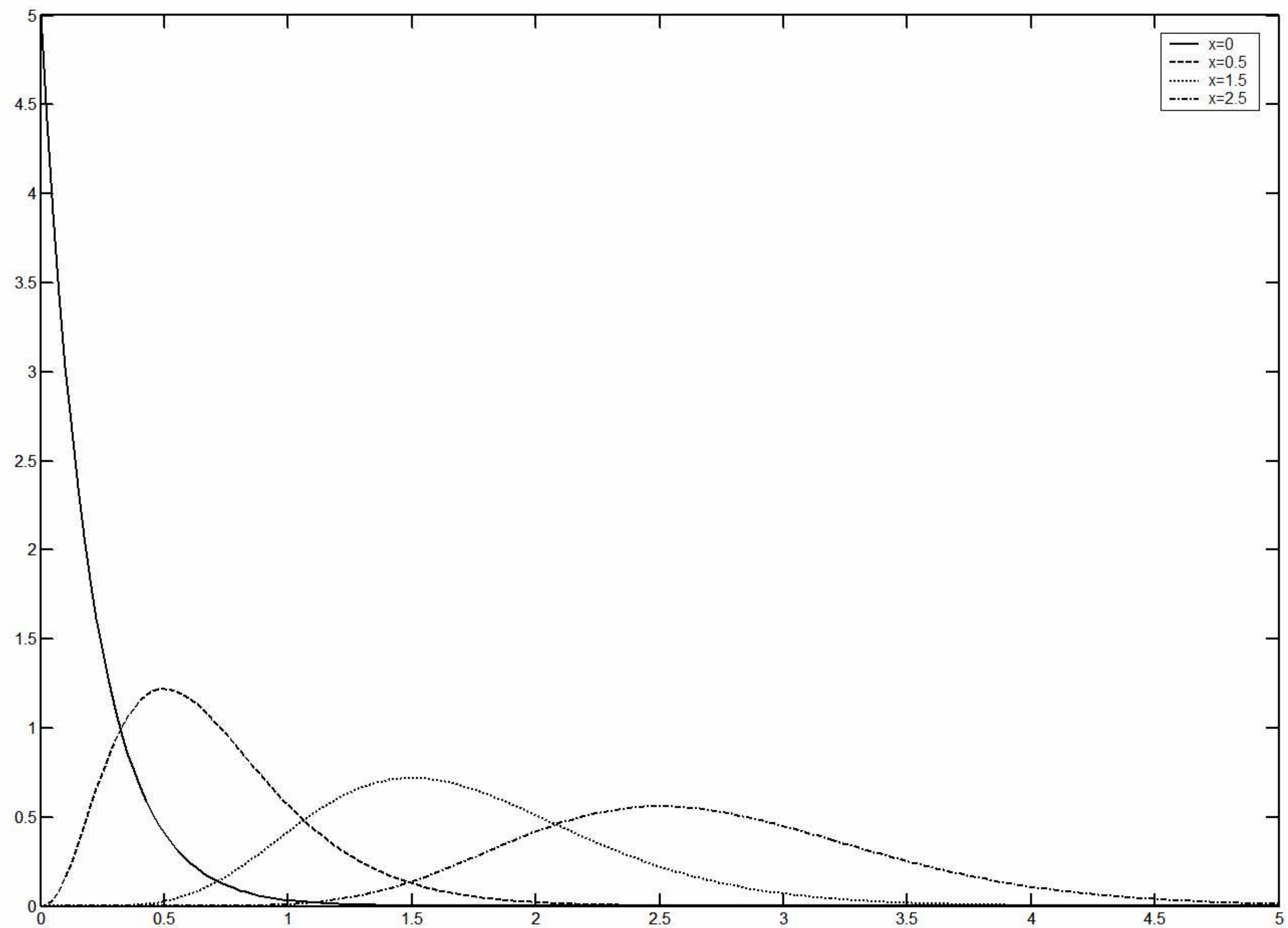

Figure 1: The gamma kernel function for different x values.

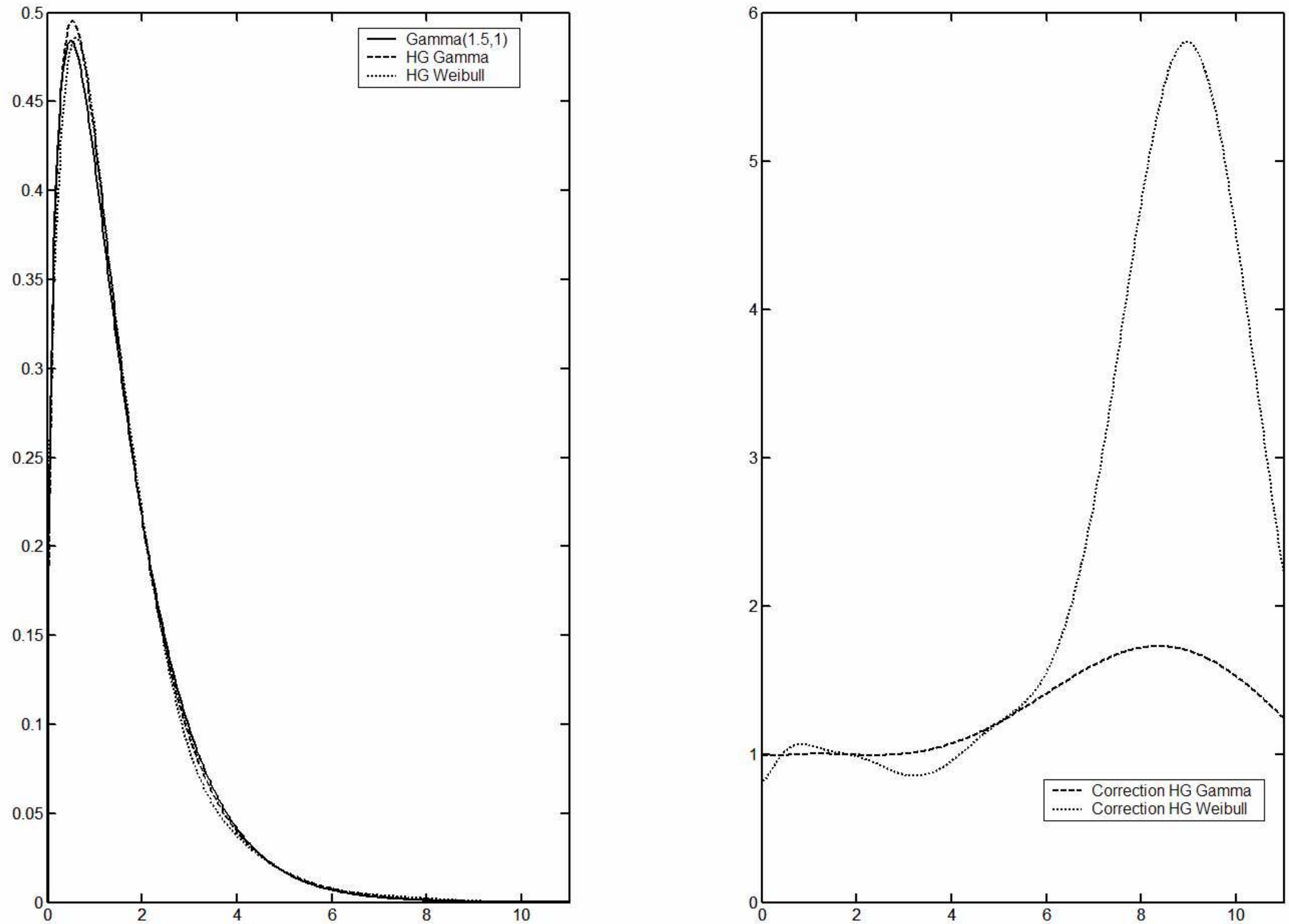

Figure 2: Density estimates and correction function for the H&G estimator with gamma and Weibull start when the true density is G(1.5,1).

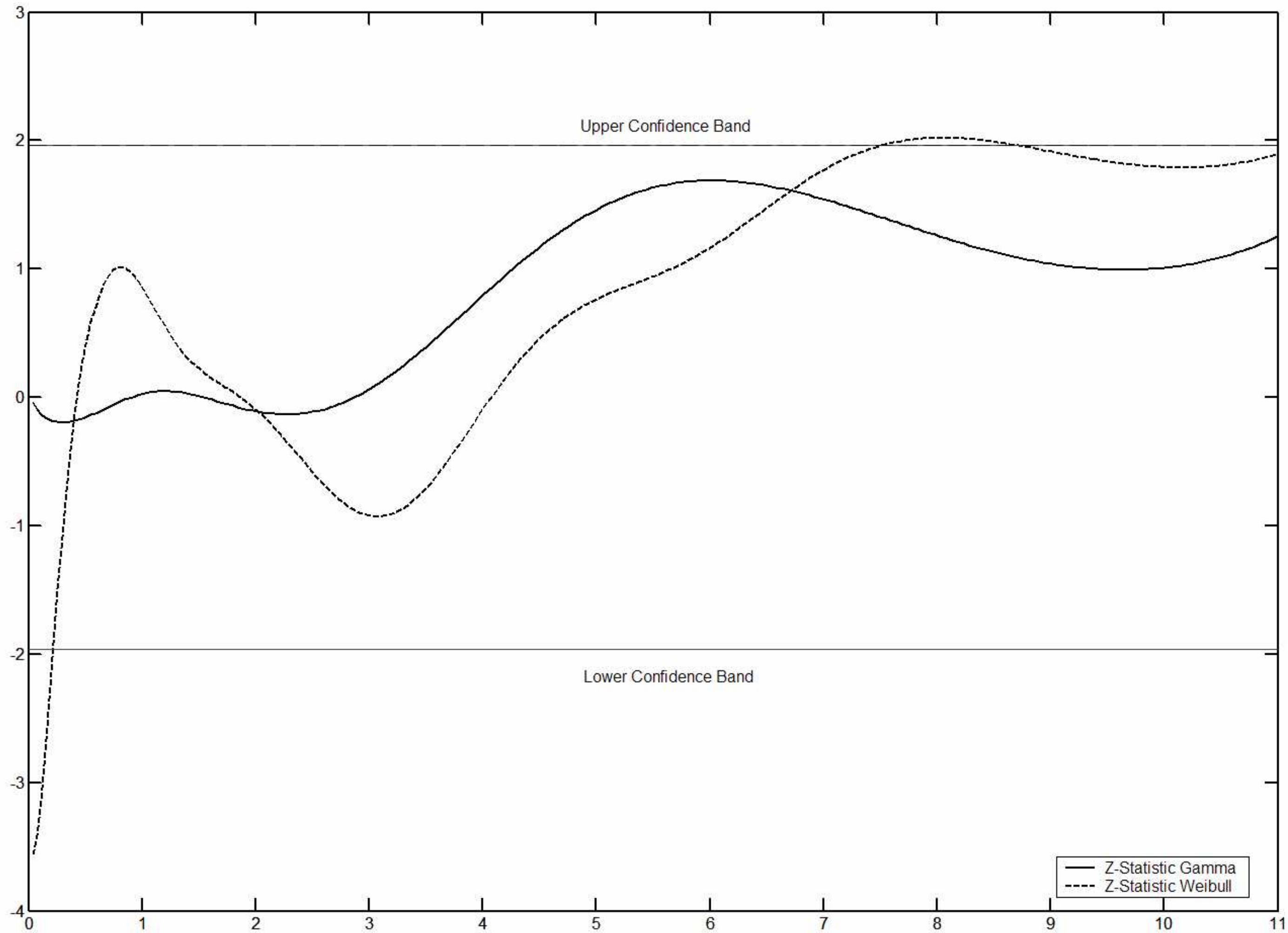

Figure 3: The Z-statistics associated to the examples in Figure 2.

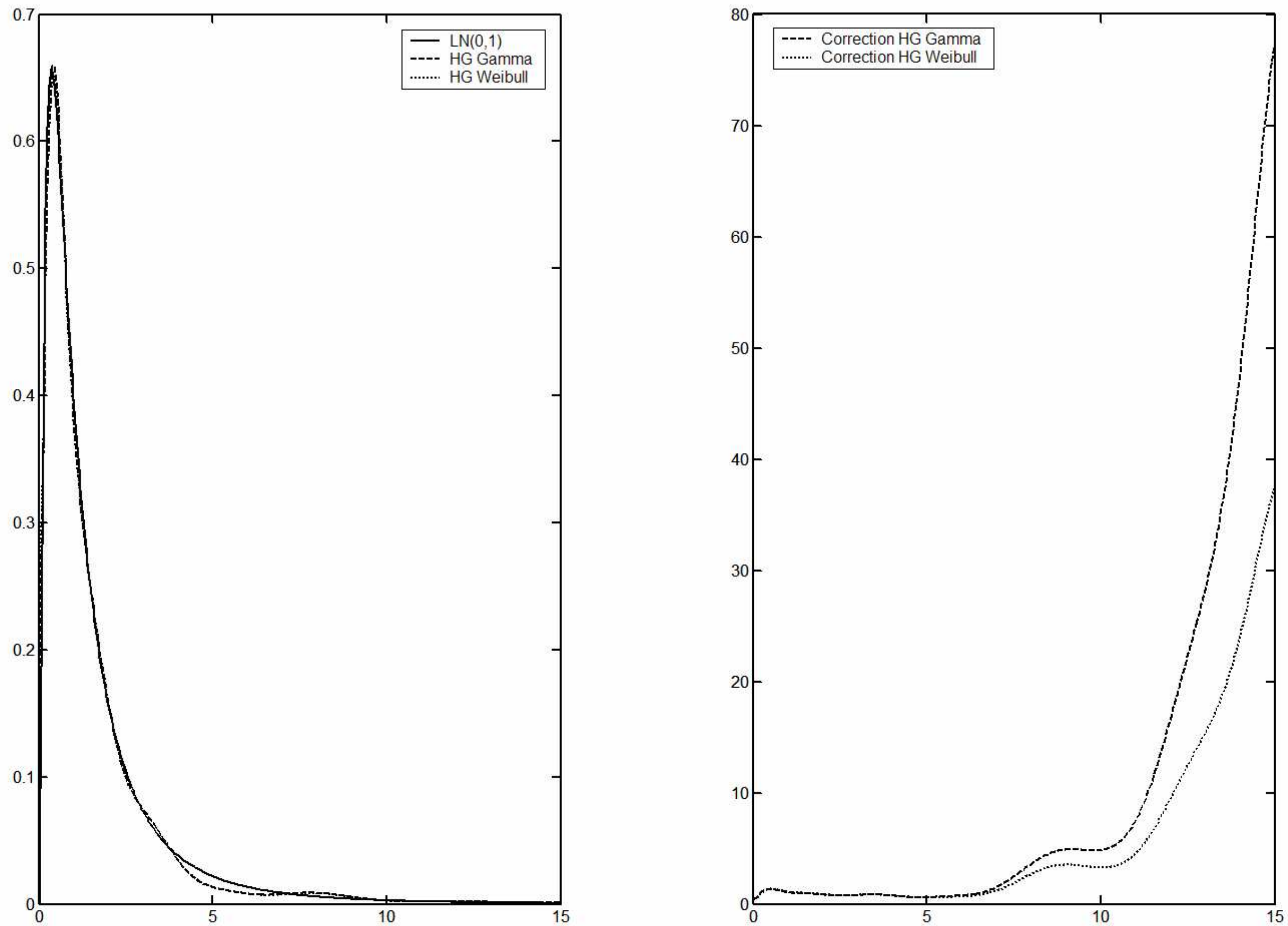

Figure 4: Density estimate and correction factor for the H&G estimator with a gamma start when the true density is LN(0,1).

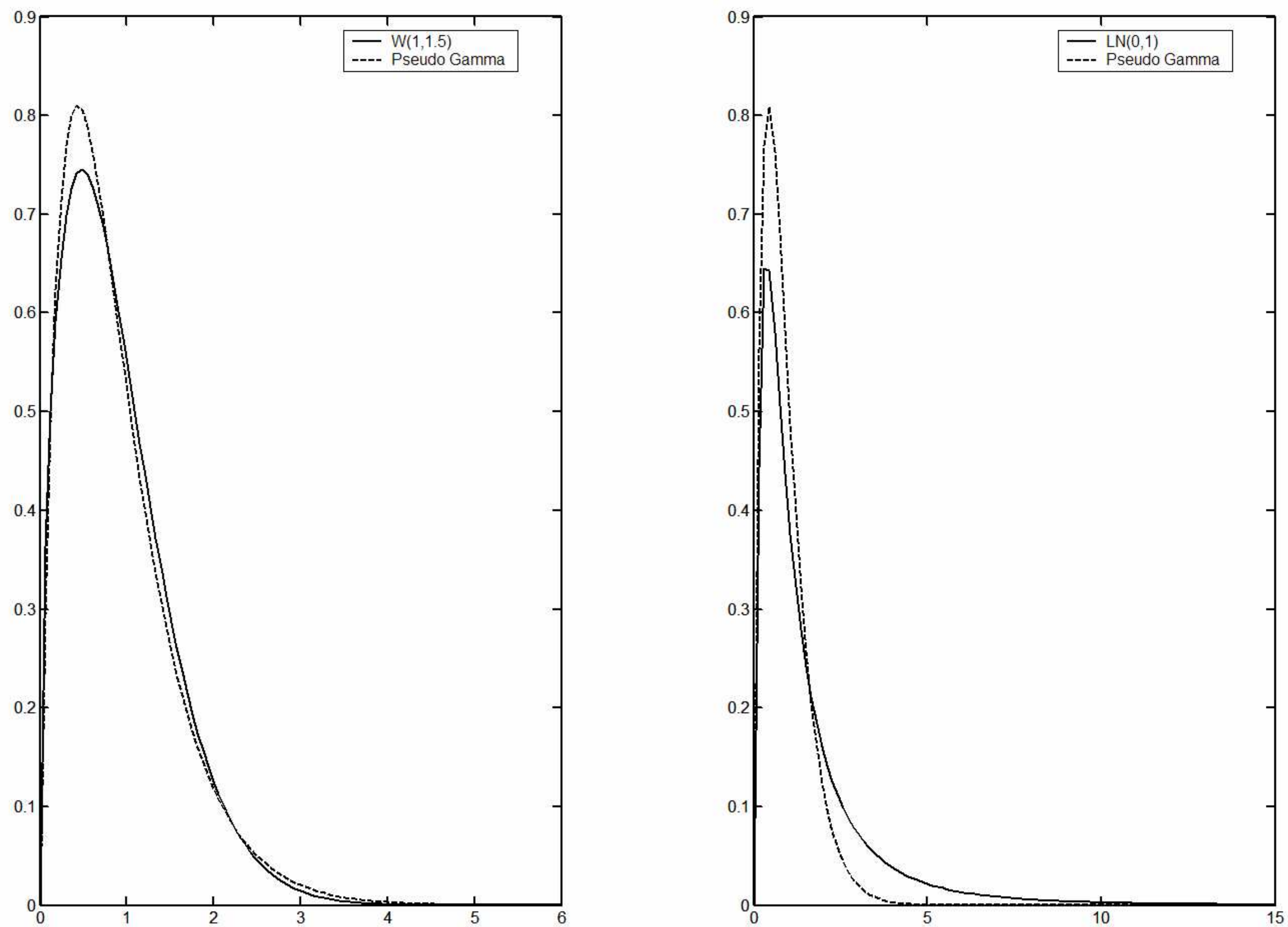

Figure 5: Pseudo gamma densities for the LN(0,1) and W(1,1.5).

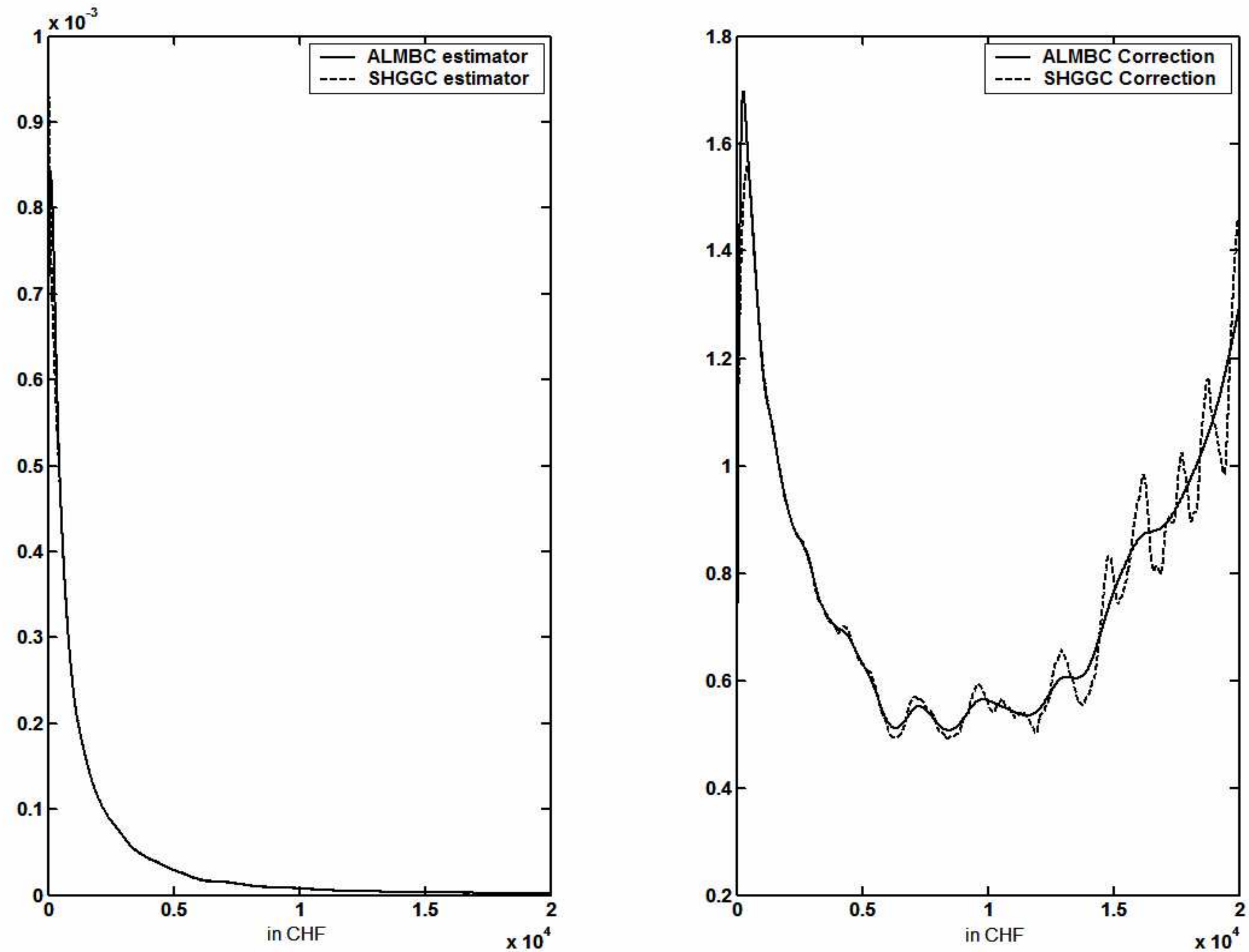

Figure 6: Loss distribution and correction factor for all clients in canton of Zurich.

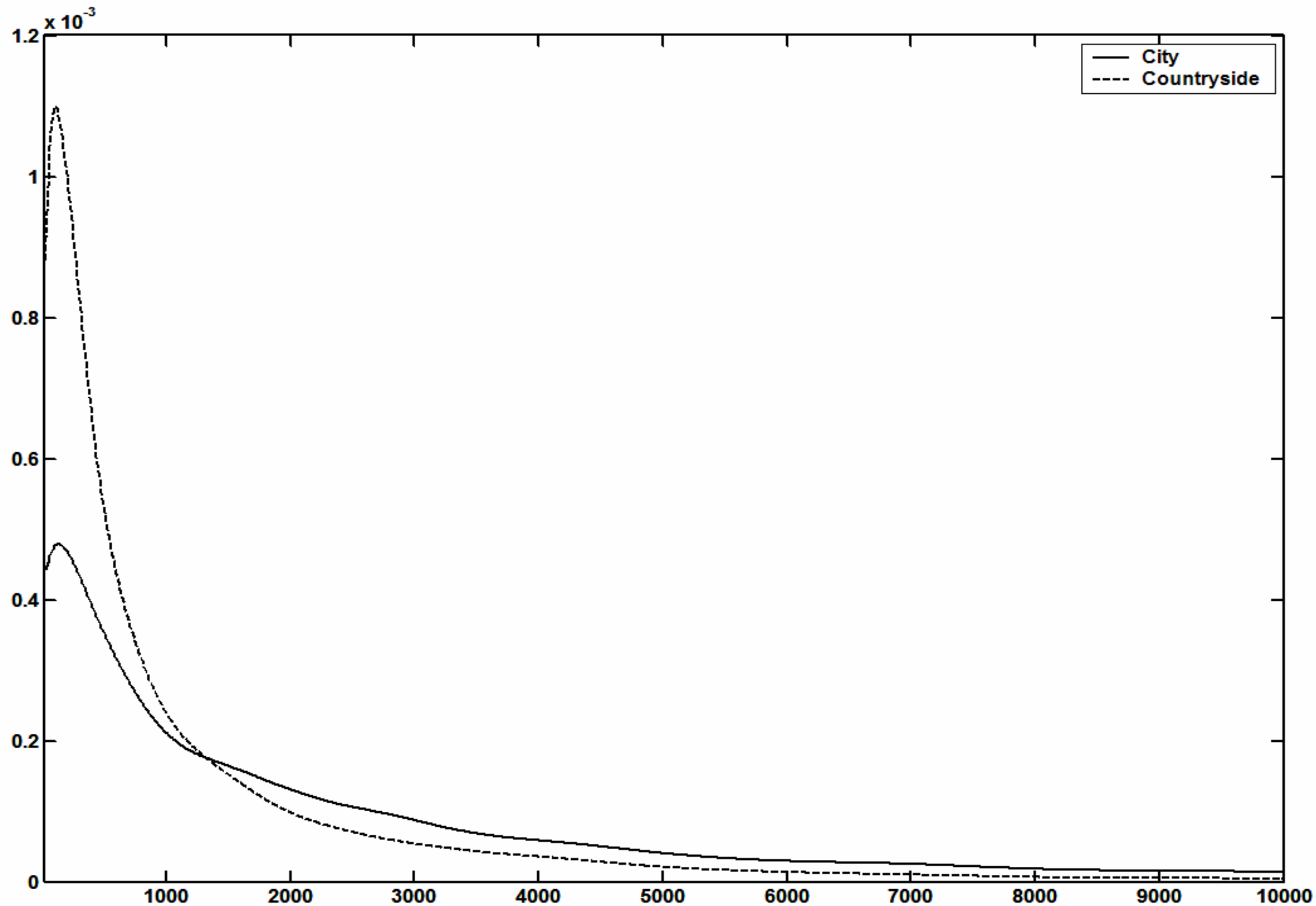

Figure 7: Loss distribution for Zurich City and countryside clients.

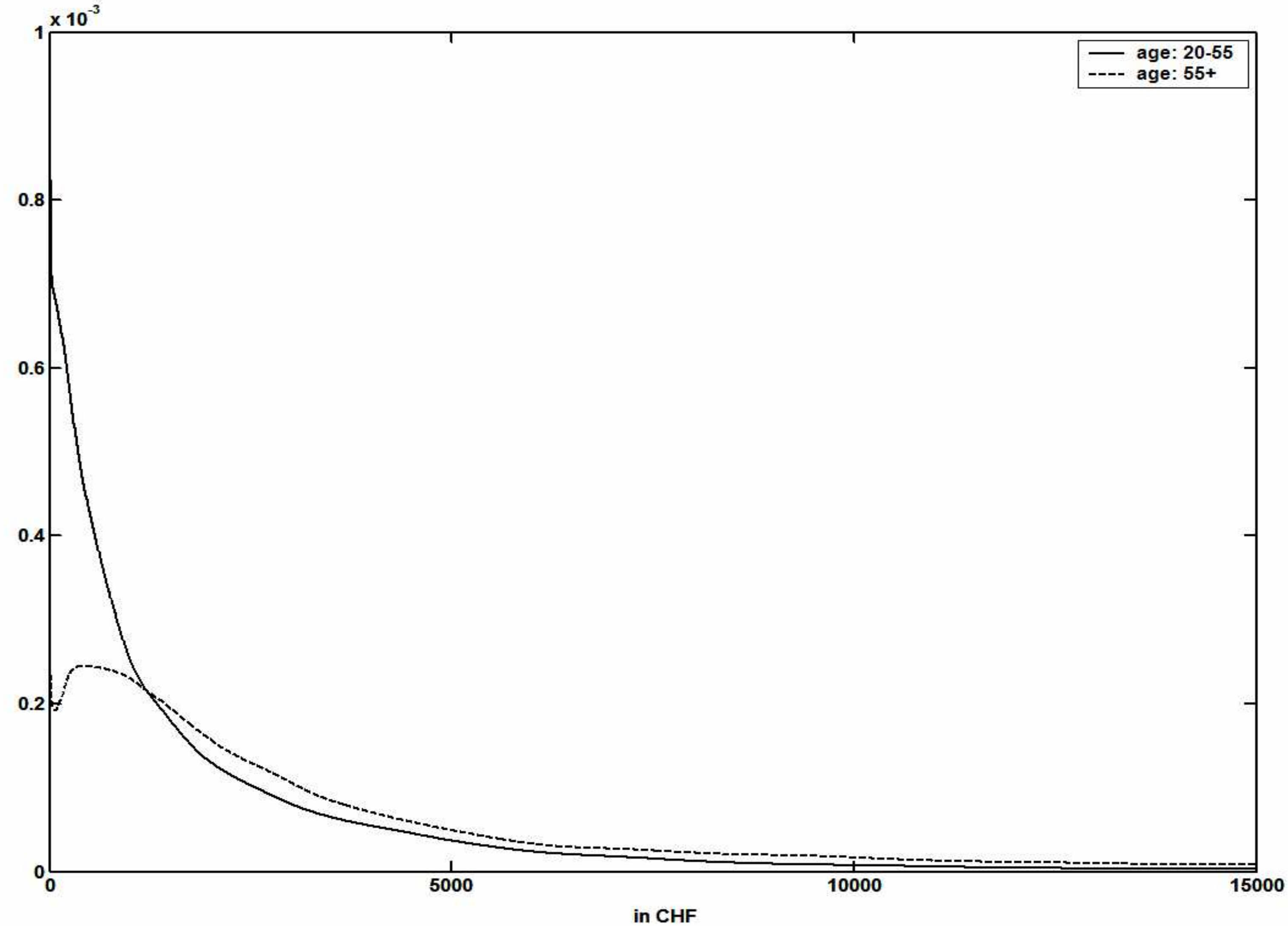

Figure 8: Loss distribution for clients with different age structure

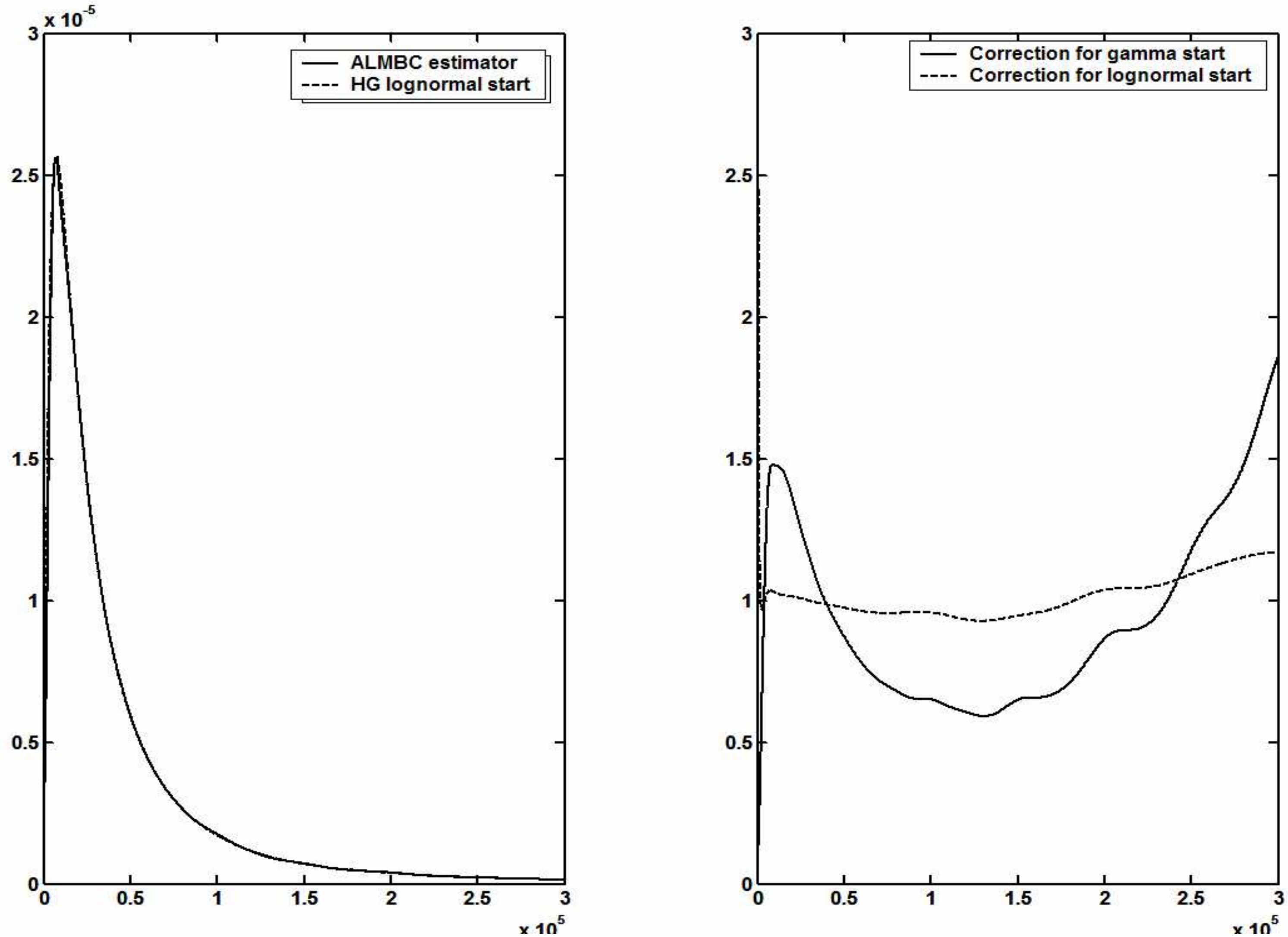

Figure 9: Brazilian income distribution and correction factors.

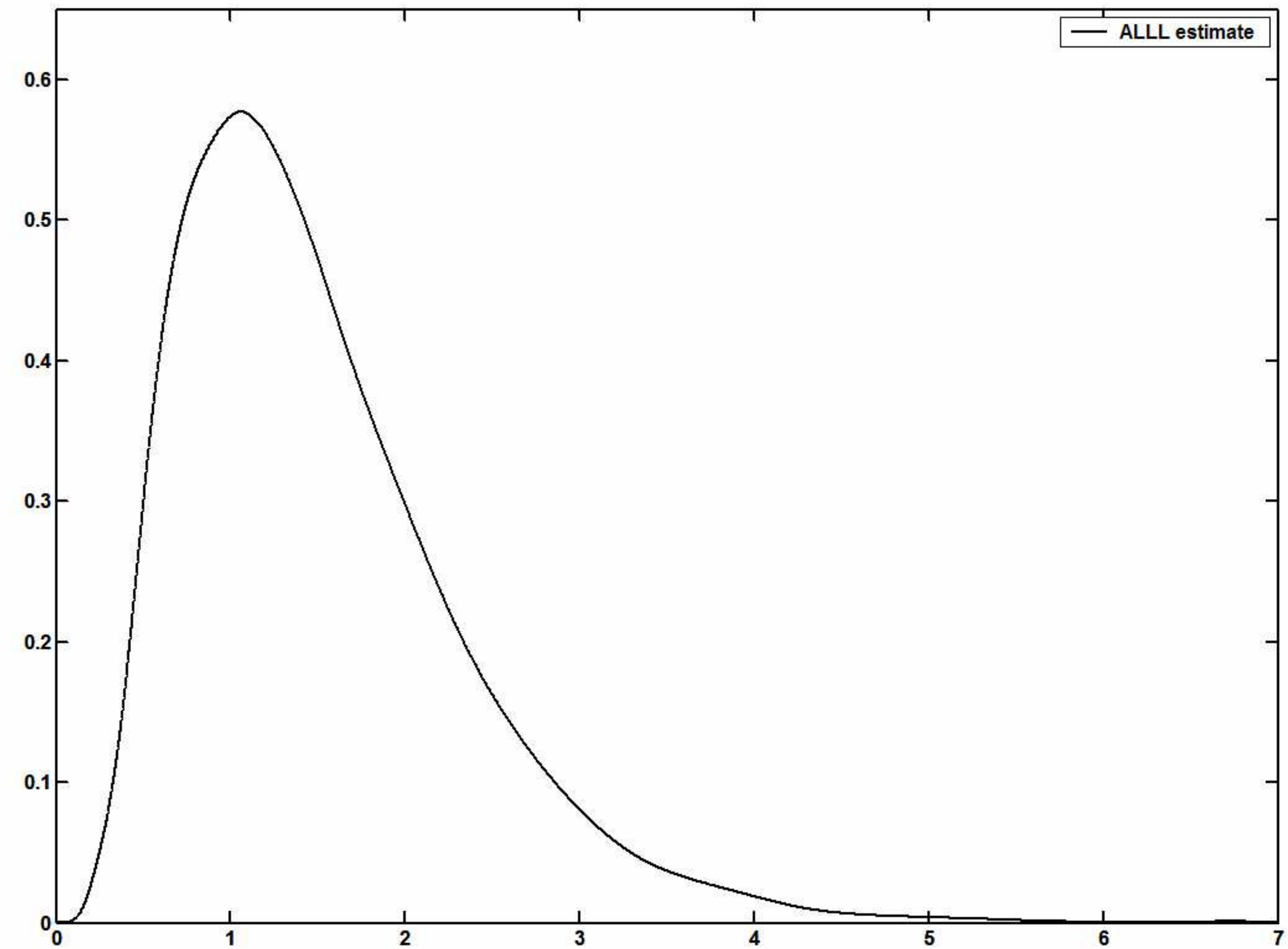

Figure 10: Density of the bootstraped test statistic values.

Table 1: MISE

Absolute Values	G1	AHGG	SHGG	SHGGC	ALMBC	ALLL
Gamma(1.5,1)						
n=200	0.00517	0.00181	0.00658	0.00216	0.00190	0.00404
n=500	0.00268	0.00071	0.00396	0.00083	0.00075	0.00205
n=1000	0.00173	0.00038	0.00267	0.00044	0.00039	0.00130
Weibull(1,1.5)						
n=200	0.00782	0.00412	0.00793	0.00561	0.00417	0.00625
n=500	0.00414	0.00208	0.00490	0.00370	0.00214	0.00328
n=1000	0.00251	0.00127	0.00341	0.00279	0.00130	0.00199
Lognormal (0,1)						
n=200	0.00790	0.00733	0.01039	0.00924	0.00586	0.00631
n=500	0.00394	0.00367	0.00561	0.00486	0.00294	0.00313
n=1000	0.00240	0.00225	0.00356	0.00303	0.00184	0.00194
Relative to G1						
Gamma(1.5,1)						
n=200	1.00	0.35	1.27	0.42	0.37	0.78
n=500	1.00	0.27	1.48	0.31	0.28	0.76
n=1000	1.00	0.22	1.54	0.25	0.23	0.75
Weibull(1,1.5)						
n=200	1.00	0.53	1.01	0.72	0.53	0.80
n=500	1.00	0.50	1.18	0.90	0.52	0.79
n=1000	1.00	0.51	1.36	1.11	0.52	0.79
Lognormal (0,1)						
n=200	1.00	0.93	1.31	1.17	0.74	0.80
n=500	1.00	0.93	1.42	1.23	0.74	0.79
n=1000	1.00	0.94	1.48	1.26	0.77	0.81

Table 2: WISE

Absolute Values	G1	AHGG	SHGG	SHGGC	ALMBC	ALLL
Gamma(1.5,1)						
n=200	0.00902	0.00380	0.01897	0.00580	0.00355	0.00686
n=500	0.00481	0.00150	0.01107	0.00238	0.00141	0.00352
n=1000	0.00295	0.00083	0.00712	0.00120	0.00074	0.00216
Weibull(1,1.5)						
n=200	0.00548	0.00275	0.00704	0.00403	0.00303	0.00443
n=500	0.00277	0.00143	0.00400	0.00241	0.00149	0.00224
n=1000	0.00168	0.00088	0.00270	0.00172	0.00090	0.00138
Lognormal (0,1)						
n=200	0.01762	0.01719	0.04940	0.04472	0.01284	0.01367
n=500	0.00859	0.00824	0.02563	0.02308	0.00620	0.00656
n=1000	0.00518	0.00495	0.01604	0.01438	0.00380	0.00402
Relative to G1						
Gamma(1.5,1)						
n=200	1.00	0.42	2.10	0.64	0.39	0.76
n=500	1.00	0.31	2.30	0.49	0.29	0.73
n=1000	1.00	0.28	2.42	0.41	0.25	0.73
Weibull(1,1.5)						
n=200	1.00	0.50	1.29	0.73	0.55	0.81
n=500	1.00	0.52	1.44	0.87	0.54	0.81
n=1000	1.00	0.52	1.60	1.02	0.53	0.82
Lognormal (0,1)						
n=200	1.00	0.98	2.80	2.54	0.73	0.78
n=500	1.00	0.96	2.98	2.69	0.72	0.76
n=1000	1.00	0.95	3.09	2.77	0.73	0.77

Table 3: Sample Statistics for Health Insurance Data

	Wh. Sample	Men	Women	Age 20-55	Age 55+	Zurich City	Countryside
Sample Size	42'722.00	17'478.00	25'244.00	15'579.00	12'098.00	8'737.00	8'670.00
Mean payment	2'971.30	2'582.90	3'240.30	2'984.50	5'697.20	4'960.60	1'966.70
Stand. Dev.	6'671.70	6'488.60	6'782.70	6'542.10	9'119.70	9'221.60	5'017.40
Skewness	7.35	9.24	6.23	7.94	5.09	5.85	9.50
Kurtosis	102.94	152.85	74.74	102.53	58.44	71.41	151.12
Minimum	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Maximum	202'870.00	202'870.00	169'650.00	132'890.00	202'870.00	202'870.00	126'700.00
1st quartile	320.30	278.80	359.35	403.20	1'051.80	575.40	237.25
Median	923.28	761.65	1'067.00	1'165.30	2'473.80	1'804.90	612.95
3rd quartile	2'695.80	2'193.40	3'039.20	3'014.40	6'000.10	4'879.10	1'749.50

Table 4: LSCV Bandwidth Health Data

	Whole	Age 20-55	Age 55+	City	Country
G1	0.0063	0.0116	0.0199	0.0135	0.0093
ALLL	0.0300	0.0711	0.1649	0.0873	0.0560
ALMBC	0.0443	0.1983	0.1331	0.1400	0.0975
SHGGC	0.2118	-	-	-	-

Table 5: Sample Statistics Income Data

	Income Data
Sample Size	71'523.00
Mean payment	52'183.00
Standard deviaton	90'661.00
Skewness	11.01
Kurtosis	319.32
Minimum	2.00
Maximum	5'011'000.00
1st quartile	12'165.00
Median	26'142.00
3rd quartile	57'000.00

Table 6: LSCV Bandwidth Income Data

G1	0.0233
ALLL	0.0835
ALMBC	0.0856
AHGLOGN	0.2168